Features of rearrangements of human immunoglobulin light chains

Tahir, Romeeza

Awarding institution:
King's College London

Features of rearrangements of human immunoglobulin light chains

By

Romeeza Tahir

Supervised by

Professor Jo Spencer

Peter Gorer Department of Immunobiology

School of Medicine

King's College London

London, UK

A thesis submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy in Immunology at King's College London, University of London

2015

Declaration

I declare that the work described in this thesis has been personally prepared and all experiments are personally carried out. Sources of information used during the production of this thesis are acknowledged by means of a reference.

Romeeza Tahir

April 2015

Acknowledgements

First of all I am very grateful to my respected supervisor, Dr. Jo Spencer, for not only giving me an exciting opportunity to work in her laboratory but also for sharing many mind stimulating ideas throughout this endeavour to achieve our goals. She provided full guidance and support with patience and also encouraged me to think from many perspectives scientifically.

I would like to thank all members from Dr. Jo Spencer lab (Dr. Louise Fraser, Dr. Yuan Zhao, Dr. Jessica Thomas, Dr. Pamela Lutalo, Dr. Manu Shankar-Hari, Dr. Verena Hehle and Dr. Dia Chavele) and Dr. Deborah Dunn-Walters lab (Dr. Victoria Martin, Dr. Joselli Silva O'Hare, Dr. Tihomir Dodev, Catherine Townsend, and Dr. Yu-Chang Wu) for providing an environment conducive to my progress.

I would also like to thank my thesis committee (Dr. Valerie Corrigall, Dr. Susan John, Dr. Deborah Dunn-Walters, Dr. Paul Fields and Dr. Paul Blair) for their input to make sure smooth progression of project.

I would like to extend my gratitude to Dr. Deborah Dunn-Walters and Dr. David Kipling (Cardiff University) for their great support in high throughput sequencing experiments and data analysis.

I'm also grateful to Valerie Wicksey, Dr. Efstathios Theodoridis, Dr. Stelios Papaioannou, Ian Williams and Terry Bryan for offering official and technical help whenever I needed.

Abstract

B cells recognize a wide range of antigens by their specific surface B cell receptor. BCRs are immunoglobulin polypeptides and made up of heavy chain (IGH) and light chain. Light chains are of two types: kappa (IGK) or lambda (IGL). The variable regions of immunoglobulin chains are generated by somatic rearrangement of V, D and J segments to generate a diverse repertoire of BCRs. The aim of this thesis is to analyze features of immunoglobulin gene rearrangements in the context of heavy and light chain pairings and light chain gene expression.

Mature naïve B cells generated in blood or tissues have a ratio of IGK:IGL expression of approximately 2:1. In contrast IgA plasma cells exhibit IGK:IGL ratio of approximately 1:1. Such a bias could be acquired by antigen selection or by changes in light chain gene rearrangements such as light chain revision. In order to better understand this, *IGL* genes were PCR amplified, cloned and sequenced from the DNA of transitional, naïve and IgA expressing B cells. Unusual biases in genetic reading frame were observed in the IgA expressing cells of the most commonly used IGL families. This was further investigated using a high throughput sequencing method that identified biases in the *IGL* genes of IgA with preferential selection of *IGLV*2-14 in IgA subsets (Chapter 3).

In the bone marrow, *IGL* gene rearrangement starts only if the rearrangement at *IGK* locus is non-functional. However, approximately 25% to 30% of IGL expressing cells have previously undergone apparently functional *IGK* gene rearrangement, but for some reasons these productive rearrangements were not expressed. In order to better understand the regulation of the *IGK* locus, *IGK* gene rearrangements were sequenced from naïve B cells expressing either IGK or IGL. Sequences were divided into productive and non-productive according to the characteristics of the junction. The actual expressed genes in cDNA were also amplified from IGK expressing B cells. Overall results identified selection and expression biases operating at the time of establishment of *IGK* repertoire (Chapter 4).

During B cell development, *IGH* rearrangements that pair better with surrogate light chains are selected at the pro-B cell stage. Therefore, it was decided to investigate if there is any possibility that later during B cell development, rearranged light chains could show

preferences for particular *IGH* gene segments. If so, this will help in identifying additional factors shaping the *IGH* repertoire of naïve B cells. Rearranged *IGH* were sequenced from naïve B cells expressing either IGK or IGL. The relative usage of *IGHV*, *IGHD* and *IGHJ* and CDR-3 characteristics were compared. It was found that there is no preference of IGH to be associated with either IGK or IGL (Chapter 5).

This thesis has identified some unusual features of light chain gene rearrangements. The relevance this data to B cell immunology will be discussed.

Table of contents

List of figures

List of tables

List of abbreviations

ADCC	Antibody dependent cell mediated cytotoxicity
AID	Activation induced deaminase enzyme
APRIL	A proliferation inducing ligand
BACH2	Basic leucine zipper protein
BAFF	B cell activating factor of TNF family
BAFF-R	BAFF receptor
BCL6	B-cell lymphoma 6 protein
BCMA	B cell maturation antigen
BCR	B cell receptor
BTK	Bruton's tyrosine kinase
CCR7	Chemokine receptor
C_2	Cysteine
CD	Cluster of differentiation
CD40L	Cluster of differentiation 40 ligand
CDR	Complementary determining region
CLP	Common lymphoid progenitor
COOH	Carboxyl terminal
CSR	Class switch recombination
CTD	C-terminal domain
CVID	Common variable immunodeficiency
D	Diversity exon gene segment
DCs	Dendritic cells
DMSO	Dimethyl sulfoxide
DNA-PK	DNA dependent protein kinase
dsDNA	Double stranded DNA
DZ	Dark zone
Erk	Extracellular signal regulated kinase
FAE	Follicle associated epithelium
FCS	Fetal calf serum

FDC	Follicular dendritic cell
FMO	Fluorescence minus one
Foxp3	Forkhead box P3 protein
FWR	Frame work region
GALT	Gut associated lymphoid tissue
GC	Germinal center
GPCR	G-protein coupled receptor
H3K4me3	Trimethylated histone H3 lysine 4
HEVs	High endothelial venules
HSC	Haematopoietic stem cell
ICAM	Intracellular adhesion molecule
IEL	Intraepithelial lymphocyte
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IGH	Immunoglobulin heavy chain locus
IGH	Immunoglobulin heavy chain protein
IGHC	Immunoglobulin heavy chain constant exon gene
IGHC	Immunoglobulin heavy chain constant exon protein
IGHD	Immunoglobulin heavy chain diversity exon gene
IGHD	Immunoglobulin heavy chain diversity exon protein
IGHJ	Immunoglobulin heavy chain joining exon gene
IGHJ	Immunoglobulin heavy chain joining exon protein
IGHV	Immunoglobulin heavy chain variable exon gene
IGHV	Immunoglobulin heavy chain variable exon protein
IGK	Immunoglobulin kappa light chain locus
IGK	Immunoglobulin kappa light chain protein
IGKC	Immunoglobulin kappa light chain constant exon gene
IGKC	Immunoglobulin kappa light chain constant exon protein

IGKJ	Immunoglobulin kappa light chain joining exon gene
IGKJ	Immunoglobulin kappa light chain joining exon protein
IGKV	Immunoglobulin kappa light chain variable exon gene
IGKV	Immunoglobulin kappa light chain variable exon protein
IGL	Immunoglobulin lambda light chain locus
IGL	Immunoglobulin lambda light chain protein
IGLC	Immunoglobulin lambda light chain constant exon gene
IGLC	Immunoglobulin lambda light chain constant exon protein
IGLJ	Immunoglobulin lambda light chain joining exon gene
IGLJ	Immunoglobulin lambda light chain joining exon protein
IGLV	Immunoglobulin lambda light chain variable exon gene
IGLV	Immunoglobulin lambda light chain variable exon protein
IgM	Immunoglobulin M
IL	Interleukin
ILF	Isolated lymphoid follicle
iRSS	Intronic recombination signal sequence
ITAM	Immunoreceptor tyrosine based activation motif
J chain	Joining chain
kDa	Kilo Dalton
KDE	Kappa deleting element
LN	Lymph node
LP	Lamina propria
LZ	Light zone
M cell	Microfold cell
MAdCAM-1	Mucosal vascular addressin cell adhesion molecule-1
MALT	Mucosa associated lymphoid tissue
MHC	Major histocompatibility complex
MID	Multiplex identifier
MLN	Mesenteric lymph node
mRNA	Messenger ribonucleic acid

N nucleotide	Non templated nucleotide
NBD	Nonamer binding domain
NH_2	Amino terminal
NHEJ	Non homologous end joining
NK cell	Natural killer cells
OH^-	Hydroxyl group
ORF	Open reading frame
P nucleotide	Palindromic nucleotide
PAMP	Pathogen associated molecular pattern
Pax5	Paired box 5 protein
PD-1	programmed cell death-1
PHD	Plant homeodomain
PP	Peyer's patch
pIgR	Polymeric immunoglobulin receptor
Pre B cell	Precursor B cell
Pre BCR	Pre B cell receptor
Pro B cell	Progenitor B cell
RA	Retinoic acid
RA	Rheumatoid arthritis
RAG-1	Recombination activating gene-1 enzyme
RAG-2	Recombination activating gene-2 enzyme
RALDH	Retinaldehyde dehydrogenase enzyme
RBC	Red blood cell
REC	Recombination excision circle
ROS	Reactive oxygen species
RSS	Recombination signal sequence
S region	Switch region
SHM	Somatic hypermutation
SLE	Systemic lupus erythematousus
SLyRT	Sort lysis reverse transcriptase

ssDNA	Single stranded DNA
STAT	Signal transducer and activator of transcription
Syk	Spleen tyrosine kinase
TACI	Transmembrane activator
TCR	T cell receptor
TD	T dependent
TdT	Terminal deoxynucleotidyl transferase enzyme
Tfh	T follicular helper cell
TGF-β	Transforming growth factor beta
TID	T independent
TNF	Tumour necrosis factor
V	Variable exon gene segment
VCAM	Vascular cell adhesion molecule
YY1	Yin Yang 1
ZDD	Zinc binding dimerization domain
Zn	Zinc
γδ T cell	Gamma delta T cell
μHC	μ heavy chain

Chapter 1

Introduction

1. Introduction

The immune system has evolved to protect living organisms from the damaging effects of infectious agents, harmful substances such as toxins and altered self tissues such as tumours. Abnormal functioning of the immune system results in a wide range of immunopathologies. The first immune barrier is provided by the innate immune system. Innate immune mechanisms mainly include recognition of pathogen associated molecular patterns (PAMPs) by a fixed repertoire of pattern recognition receptors (PRRs), complement activation, phagocyte recruitment and production of antimicrobial proteins besides physical barriers provided by the epithelium and mucosal layers. Adaptive immunity that comes into play later generates antigen specific and long lasting immune responses. It culminates in the activation of B and T lymphocytes after recognition of antigens via B cell receptors (BCRs) and T cell receptors (TCRs) respectively. The effector functions of B cells are mediated by secretory proteins called antibodies. Antibodies are produced by differentiated B cells (plasma cells) to eliminate the antigen from the system. Antibodies are the secreted form of the BCRs. These soluble proteins are secreted in large quantities.

1.1 The structure of human immunoglobulin molecules

Immunoglobulins are members of the immunoglobulin superfamily of proteins. When they are expressed on the surface of the cells they form the BCR and when they are secreted they may be referred to as antibodies. The immunoglobulin molecule is made up of two different types of protein chains: heavy chain proteins (IGH) which are covalently linked by disulphide bonds and light chains linked to IGH by disulphide bonds (Figures 1-1 and 1-2) (Edelman and Benacerraf 1962, Sela-Culang, Kunik et al. 2013). Light chains are of two types (or isotypes): kappa (IGK) and lambda (IGL). Each immunoglobulin molecule is comprised of either IGK or IGL chain associated with the IGH chain on the cell surface generally (Neuberger, Caskey et al. 1989). The ratios of IGK to IGL vary in different species. In mice, the average IGK:IGL ratio is 20:1, whereas in humans it is 2:1 (Ghia, Gratwohl et al. 1995).

The amino ($-NH_2$) terminal of the immunoglobulin molecule is highly variable in both IGH and light chains called VH and VL respectively. After pairing of IGH chain with either IGK or IGL

chain, variable domains (VH and VL) together constitute the antigen binding sites. Within each variable domain, there are 6 hypervariable loops (3 in each) called complementary determining regions (CDR) separated by conserved β-pleated sheets called frame work regions (FWR) (Dreyer and Bennett 1965, Wu and Kabat 1970, Sela-Culang, Kunik et al. 2013). The carboxyl (-COOH) terminal of IGH constitutes constant region (CH) (Dreyer and Bennett 1965, Putnam, Liu et al. 1979). The constant region is further divided into domains (CH1 to CH3 or CH4). There are 5 different isotypes of the immunoglobulin molecule that are defined according to CH region: μ, γ, α, δ and ε. These different IGH isotypes have different effector functions and determine the class and functional properties of immunoglobulins [immunoglobulin M (IgM), immunoglobulin G (IgG), immunoglobulin A (IgA), immunoglobulin D (IgD) and immunoglobulin E (IgE)] respectively. These immunoglobulins have different distribution in tissues and are specialized to mediate specialized functions (Tables 1-1 and 1-2) (Murphy, et al., 2008).

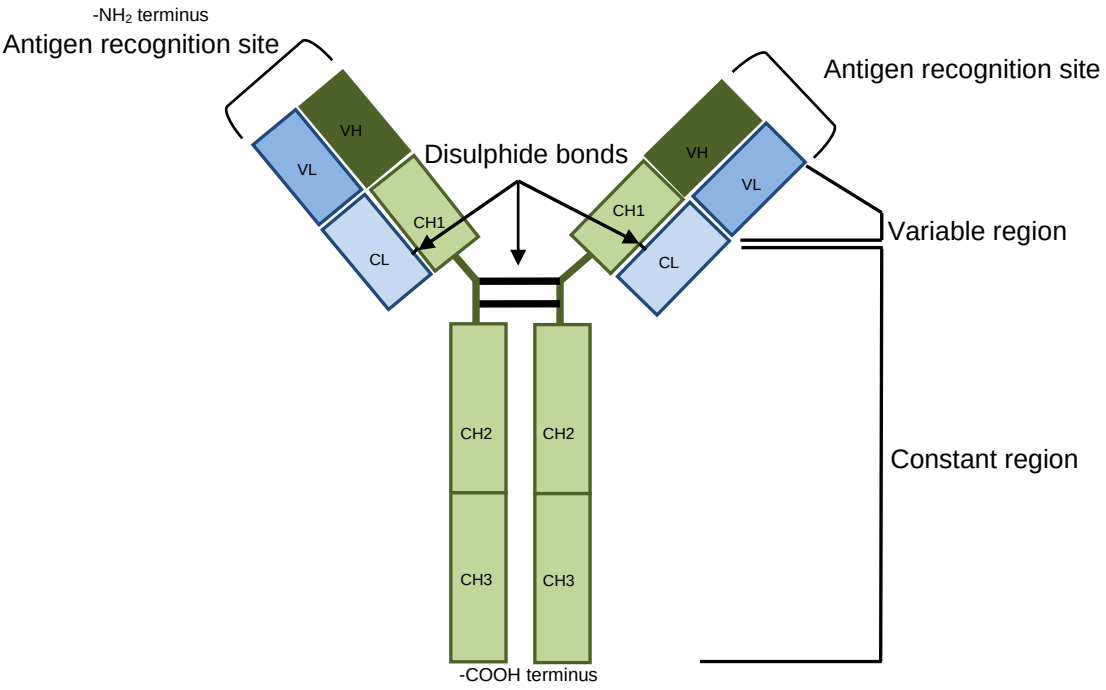


Figure 1-1: Diagrammatic representation of an immunoglobulin molecule. Each immunoglobulin is made up of two heavy chains (green) joined by disulphide bonds (black) and two light chains (blue). Each heavy chain is also linked to the light chain by the disulphide bond. Heavy and light chains are divided into variable and constant region domains.

The structure of a membrane bound BCR and its corresponding secreted antibody is identical except that membrane bound BCR has a hydrophobic transmembrane portion at the -COOH terminal and a small hydrophilic tail extending into cytoplasm. The switching from transmembrane cell receptor to secreted antibody takes place by alternative splicing of the *IGH* transcript (Rogers, Early et al. 1980). The IGH chain is also non covalently linked to the invariant signalling proteins Igα and Igβ (Schamel and Reth 2000). These signalling proteins are antigen non specific and each has a single immunoreceptor tyrosine based activation motif (ITAM) in their cytoplasmic tails that enables them to transduce signal when the BCR is ligated with its cognate antigen (Reth 1989). Igα and Igβ are connected to each other by a disulphide bond (Hombach, Lottspeich et al. 1990, Reth, Wienands et al. 1991, Schamel and Reth 2000).

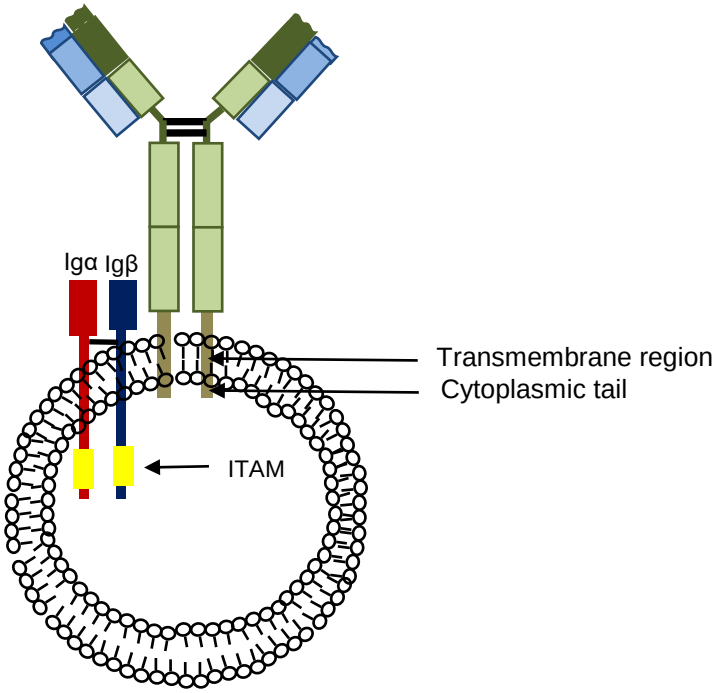


Figure 1-2: Diagrammatic representation of the BCR complex. Each immunoglobulin molecule is made up of two heavy chains (green) and two light chains (blue) joined by disulphide bonds (black). The BCR is expressed along with Igα and Igβ joined by disulphide linkage on the cell surface.

Table 1-1: The characteristic effector functions of immunoglobulins and IgG subclasses of human immunoglobulins

Effector function	IgM	IgD	IgG1	IgG2	IgG3	IgG4	IgA	IgE
Neutralization	+	-	++	++	++	++	++	-
Opsonization	+	-	+++	*	++	+	+	-
Antibody dependent cell mediated cytotoxicity	-	-	++	-	++	-	-	-
Sensitization of mast cells	-	-	-	-	-	-	-(?)	+++
Complement activation	+++	-	++	+	+++	-	+	-
Degranulation of eosinophils	-	-	-	-	-	-	++	+

Table 1-2: The distribution of various immunoglobulins and IgG subclasses of human immunoglobulins

Distribution	IgM	IgD	IgG1	IgG2	IgG3	IgG4	IgA	IgE
Transport across epithelium	+	-	-	-	-	-	+++ (dimer)	-
Transport across placenta	-	-	+++	+	++	+/-	-	-
Diffusion into extravascular sites	+/-	-	+++	+++	+++	+++	++ (monomer)	+
Mean serum level (mg/ml)	1.5	0.04	9	3	1	0.5	2.1	$3X10^{-5}$

1.2 Human gene loci encoding IGH, IGK and IGL chains

Dreyer and Bennett showed by protein sequencing that immunoglobulin molecules consist of variable and constant regions. Later, Tonegawa and Hozumi provided empirical evidence that during B cell development, functional protein of BCR is encoded by somatic gene rearrangement of widely separated variable and constant region gene segments (Dreyer and Bennett 1965, Hozumi and Tonegawa 1976). Years later, it was shown that functional variable region exon of light chain is generated by gene rearrangement of two gene segments: V (variable) and J (joining) and variable region exon of heavy chain is encoded by bringing together three gene segments: V (variable), D (diversity) and J (joining) (Bernard, Hozumi et al. 1978, Early, Huang et al. 1980).

1.2.1 Heavy chain locus

The human immunoglobulin heavy chain locus (*IGH*) is located at chromosome 14q32.33 (Figure 1-3). There are 123 to 129 *IGHV* genes: 38 to 44 functional *IGHV* genes, 4 open reading frame (ORF), 79 pseudogenes, one gene segment (*IGHV*3-11) could be in the category of either functional or pseudogene, one gene segment (*IGHV*4-61) could be either functional or ORF potentially. The potential repertoire of BCR is encoded by 38 to 46 functional *IGHV* genes belonging to 7 families defined by sequence homology (*IGHV*1 to *IGHV7*). There are 23 *IGHD*, 6 *IGHJ* and 9 *IGHC* gene segments (Lefranc, Giudicelli et al. 2015).

1.2.2 Kappa light chain locus

The human immunoglobulin kappa light chain locus (*IGK*) is located at chromosome 2p11.2 (Figure 1-4). There are 76 *IGKV* genes segments: 34 functional *IGKV* genes, 8 ORF, 30 pseudogenes, 4 genes (*IGKV*1-13, *IGKV*1-39, *IGKV*2-29 and *IGKV*3D-15) could be either functional or pseudogenes potentially. The potential BCR repertoire is encoded by 34 to 38 functional *IGKV* genes divided in to 5 families defined by sequence homology (*IGKV*1 to *IGKV*5). There are 5 *IGKJ* and 1 *IGKC* gene segments (Lefranc, Giudicelli et al. 2015).

1.2.3 Lambda light chain locus

The human immunoglobulin lambda light chain locus (*IGL)* is located at chromosome 22q11.2 (Figure 1-5). There are 73 to 74 *IGLV* gene segments: 29 to 30 functional, 5 ORF, 35 pseudogenes, 3 genes segments (*IGLV*3-9, *IGLV*3-22, and *IGLV*7-46) could be either functional or pseudogenes, one gene segment (*IGLV*1-41) could be either ORF or pseudogene. The potential BCR repertoire is encoded by 29 to 33 functional *IGLV* genes classified in to 10 families defined by sequence homology (*IGLV*1 to *IGLV*10). There are 4 *IGLJ* and 4 *IGLC* gene segments (Lefranc, Giudicelli et al. 2015).

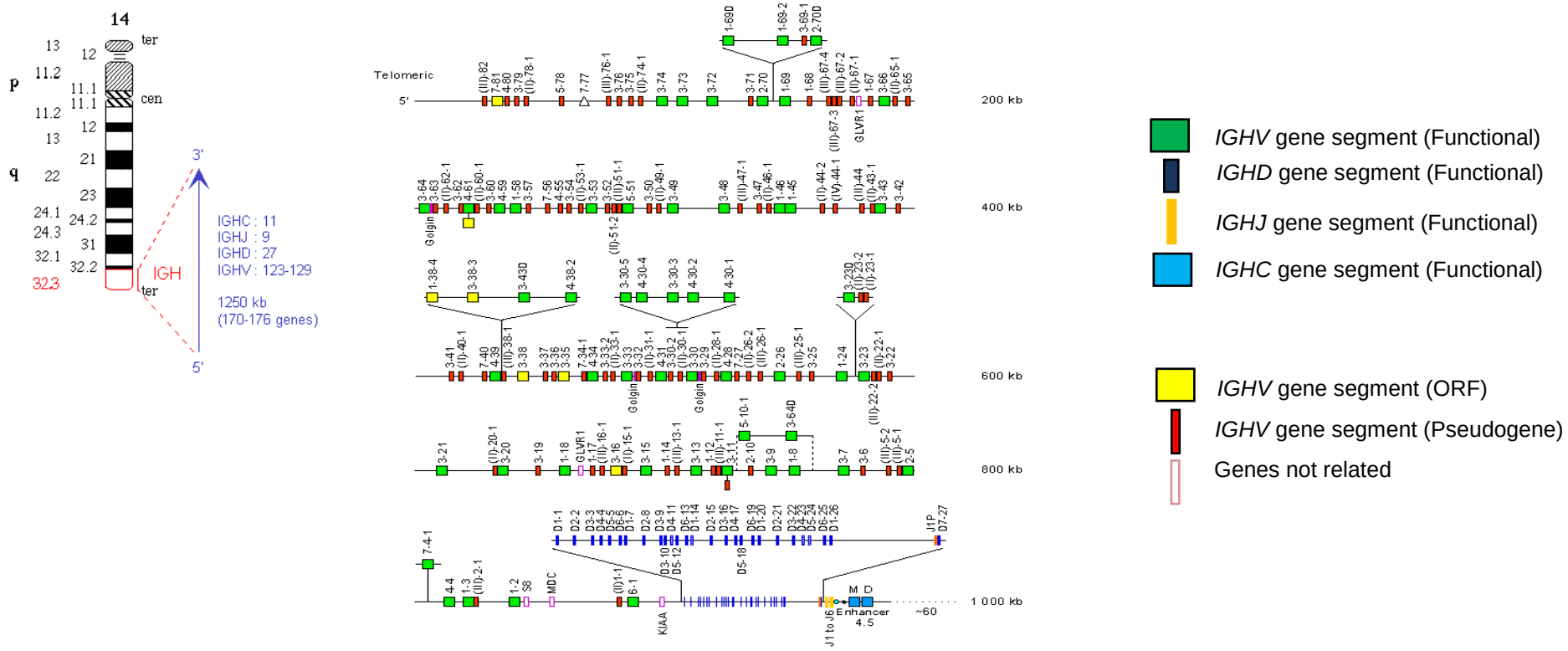


Figure 1-3: Schematic representation of human *IGH* locus located on chromosome 14q32.33. Various functional gene segments, pseudogenes and open reading frames are shown (http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=locus&species=human&group=IGH).

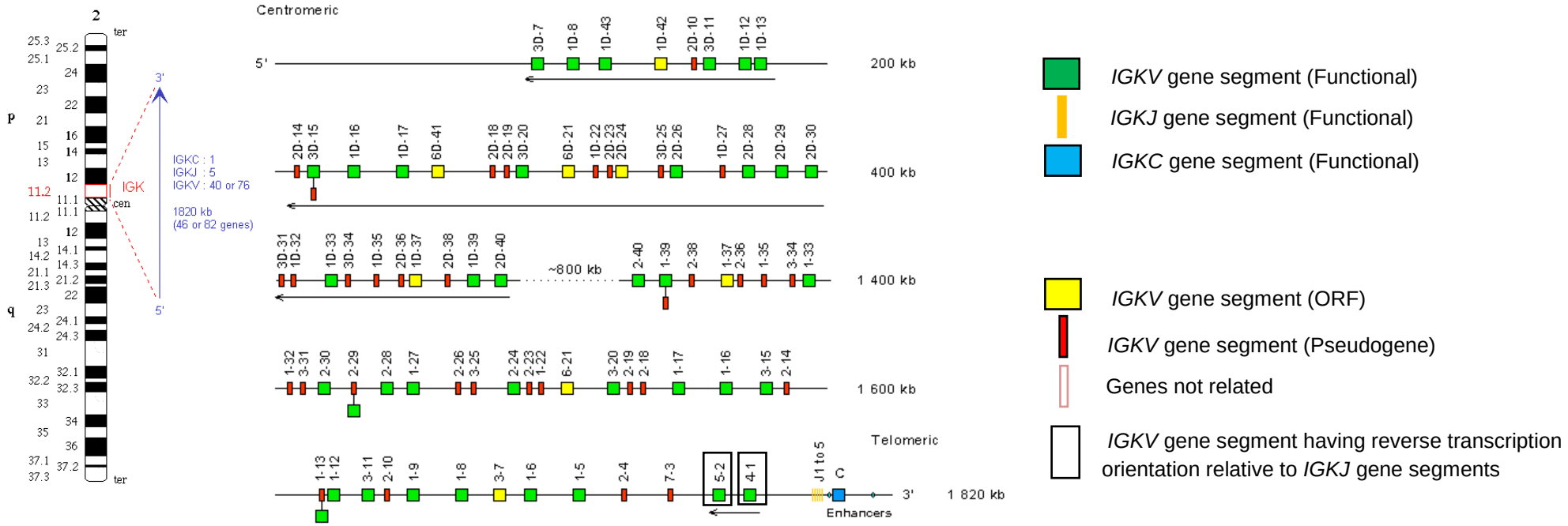


Figure 1-4: Schematic representation of human *IGK* locus located on chromosome 2p11.2. Various functional gene segments, pseudogenes and open reading frames are shown (http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=locus&species=human&group=IGK).

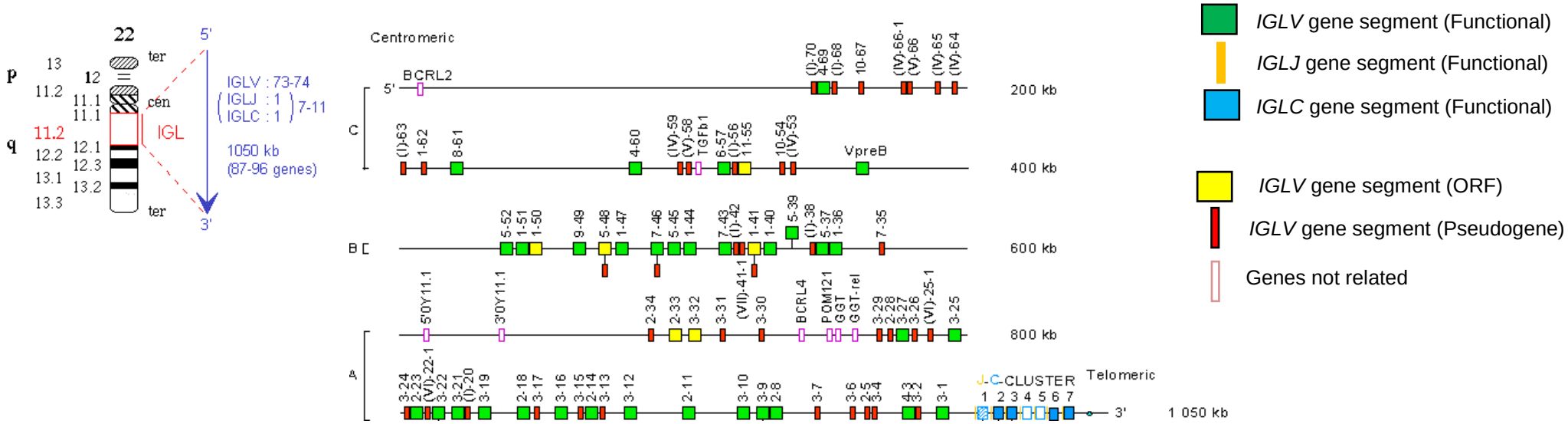


Figure 1-5: Schematic representation of human *IGL* locus located on chromosome 22q11.2. Various functional gene segments, pseudogenes and open reading frames are shown (http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=locus&species=human&group=IGL).

1.3 V(D)J gene rearrangement

V(D)J gene rearrangement is a lineage specific developmental process that assembles the variable set of exons of immunoglobulin genes to encode a functional BCR that can be expressed on the cell surface to recognize antigens (Alt, Yancopoulos et al. 1984). In order to mount an efficient adaptive immune response, it is of cardinal importance to generate a diverse repertoire of B cells (Bassing, Swat et al. 2002, Schatz and Ji 2011). The process of V(D)J gene rearrangement involves recognition and breaking of recombination signal sequences (RSSs) flanking *IGH* (*IGHV*, *IGHD*, *IGHJ)*, *IGK* (*IGKV*, *IGKJ*) and *IGL* (*IGLV* and *IGLJ)* gene segments by recombination activation gene enzyme (RAG) protein complex and subsequent ligation of gene segments resulting in a functional unit of exons that can encode a BCR to be expressed on the cell surface. This recognition, breaking and ligation occurs at specific stages of B cell development and are tightly regulated (Schatz, Oettinger et al. 1989, Oettinger, Schatz et al. 1990, Oettinger 2004).

1.3.1 RAG enzymes and recombination signal sequences

The RAG enzyme complex is a nuclear protein complex comprising of RAG-1 and RAG-2 enzymes (Schatz, Oettinger et al. 1989, Oettinger, Schatz et al. 1990). The structure of the RAG complex is well conserved among vertebrates. The expression of RAG-2 enzyme is lower during S, G2 and M phases of mitosis and RAG-1 predominantly bind accessible RSSs. Later as cells enter the G1 phase, RAG2 levels increase and it is subsequently recruited to the locus that is pre-bound with RAG-1 enzyme (Lin and Desiderio 1994, Lee and Desiderio 1999, Swanson 2004). The crystal structure of the RAG enzyme complex has been determined recently suggesting the RAG enzyme complex (230 kDa) forms a Y-shaped structure where the active site is in the middle. Various mutations in the active site are linked to immunodeficiency diseases (Kim, Lapkouski et al. 2015).

RAG-1 enzyme: Various domains of the RAG-1 enzyme are associated with different functions during V(D)J gene rearrangement as shown in Figure 1-6. The non-core domain of the RAG-1 protein at the $-NH_2$ terminal regulates the stability of many other cellular proteins, co-ordinates zinc (Zn^{++}) and thus enhances V(D)J gene rearrangement (Cortes, Ye et al. 1994, McMahan, Difilippantonio et al. 1997). The $-NH_2$ terminal contains several pairs of

conserved cysteine (C_2) residues. Zn^{++} binding dimerization domain (ZDD) spans from amino acid 265 to 380. This region mainly regulates interaction with the histone H3 (Rodgers, Bu et al. 1996). The ZDD is followed by the core region. The core region is the minimal portion of RAG protein required for V(D)J recombination (De and Rodgers 2004). The core region is further divided functionally into the following three main regions:

1. Nonamer binding domain (NBD) (from amino acid 384 to 458)
2. Central domain (from amino acid 528 to 760)
3. C-terminal domain (CTD) (from amino acid 761 to 1008)

The NBD forms a dimer and interacts with the nonamer of the RSS. The central domain has nicking activity and is also involved in making contact with the heptamer of the RSS. The CTD binds the double stranded DNA (dsDNA) and mediates contact with the coding sequence flanking the RSS and contains Zn^{++} binding sites (Difilippantonio, McMahan et al. 1996, Arbuckle, Fauss et al. 2001, Mo, Bailin et al. 2001, Schatz and Swanson 2011).

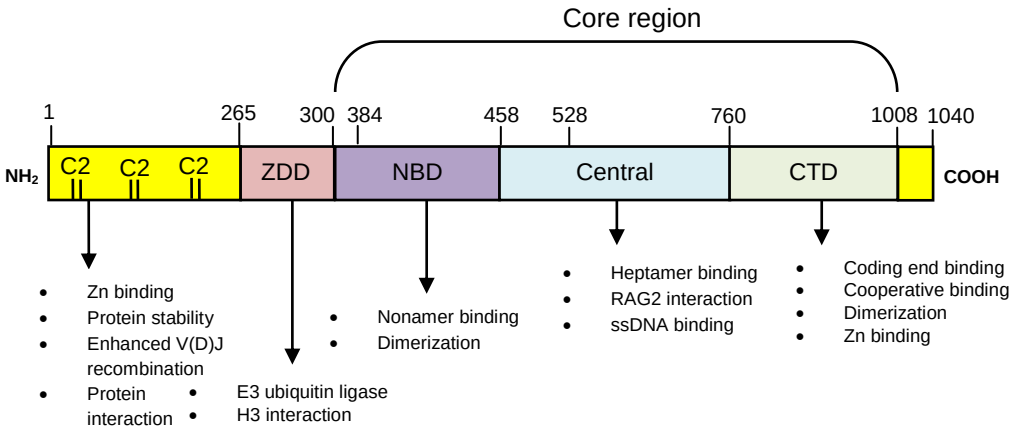


Figure 1-6: Schematic representation of RAG-1 enzyme. The position of various structural domains of protein and associated functions are indicated.

RAG-2 enzyme: Various structural domains of the RAG-2 enzyme associated with various functions are shown in Figure 1-7. The RAG-2 enzyme can be divided into following three notable regions:

1. Core region at the -NH_2 terminal (ranging from amino acid 1 to 387)
2. Flexible acidic hinge region (from amino acid 360 to 408)
3. Plant homeodomain (PHD) at the -COOH terminal (from amino acid 414 to 487)

The core region of the RAG-2 protein is critical for the cleavage of DNA. It also interacts with the RAG-1 enzyme and enhances the DNA binding affinity and specificity of the RAG complex. The PHD domain binds specifically to trimethylated histone H3 lysine 4 (H3K4me3) and enhances the catalytic activity of the RAG complex. Moreover, it directs RAG-2 to the region of active chromatin (Roldan, Fuxa et al. 2005, Liu, Subrahmanyam et al. 2007, Matthews, Kuo et al. 2007).

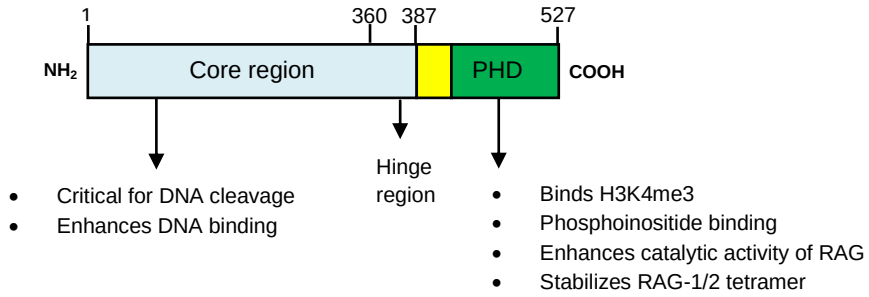


Figure 1-7: Schematic representation of RAG-2 enzyme. The position of various structural domains of proteins and associated functions are indicated.

Recombinational signal sequences: Gene segments of all immunoglobulin loci are flanked by RSSs (Hozumi and Tonegawa 1976, Rogers, Early et al. 1980). These sequences are recognized by RAG enzymes (Schatz, Oettinger et al. 1989, Oettinger, Schatz et al. 1990). Structurally each RSS consists of specific DNA elements: heptamer, nonamer and spacer DNA (Figure 1-8). The DNA sequence of the heptamer and nonomer are well conserved but the DNA sequence of spacer region is variable and it can be either

12 base pair (bp) or 23 bp which define 12-RSS or 23-RSS respectively (van Gent, Ramsden et al. 1996). The initial three nucleotides of heptamer are critical during the rearrangement process (Hesse, Lieber et al. 1989, Ramsden, McBlane et al. 1996). The NBD of the RAG-1 enzyme binds nonamer (ACAAAAACC) and mediates anchoring of RAG proteins on the DNA. The heptamer (CACAGTG) sequence enhances anchoring of the RAG complex and specifies the site of DNA cleavage. The length of spacer DNA is important for efficient V(D)J gene rearrangement as it is involved in the alignment of heptamer and nonamer. It also plays an important role in protein-DNA interactions (Swanson, Kumar et al. 2009, Schatz and Swanson 2011).

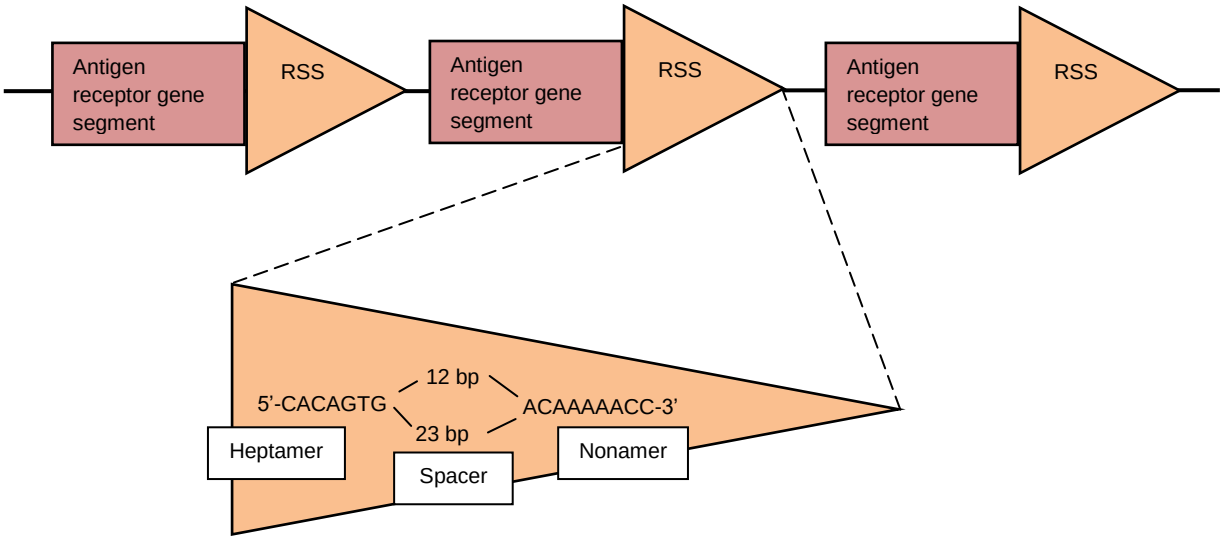


Figure 1-8: Schematic representation of RSS flanking all gene segments of immunoglobulin loci. Each RSS has conserved heptamer and nonamer along with non conserved spacer DNA.

1.3.2 Mechanism of V(D)J gene rearrangement

During V(D)J gene rearrangement the “12-23 rule” is followed that ensures specific and precise gene rearrangements. According to this rule, there is a restriction of gene rearrangement between 12-RSS and 23-RSS only (van Gent, Ramsden et al. 1996). Theoretically, the process of V(D)J gene rearrangement can be divided into two phases as described below (Fugmann, Lee et al. 2000, Gellert 2002):

1.3.2.1 Cleavage phase

RAG enzymes recognize both 12-RSS and 23-RSS and introduce nicks on the DNA strands. Nicking of the DNA results in the alignment of the two gene segments and leads to the formation of a stable multi subunit synaptic complex (Jones and Gellert 2002, Mundy, Patenge et al. 2002, Curry, Geier et al. 2005). RAG enzymes first introduce a nick on one of the DNA strands by catalyzing the hydrolysis of a phosphodiester bond adjacent to the RSS. Free 3’ hydroxyl (OH) group liberated as a result of nicking attacks on the opposite strand of the DNA and thus a double stranded break is generated. It is then followed by hairpin formation on both coding strands resulting in a blunt signal end (McBlane, van Gent et al. 1995) (Figure 1-9). RAG proteins remain bound transiently with the broken ends of the DNA that are produced after cleavage (Hiom and Gellert 1998). The coding ends dissociate from RAG proteins before the signal ends. Later there is a progression to the next phase that results in the joining of the broken DNA ends (Jones and Gellert 2002, Mundy, Patenge et al. 2002).

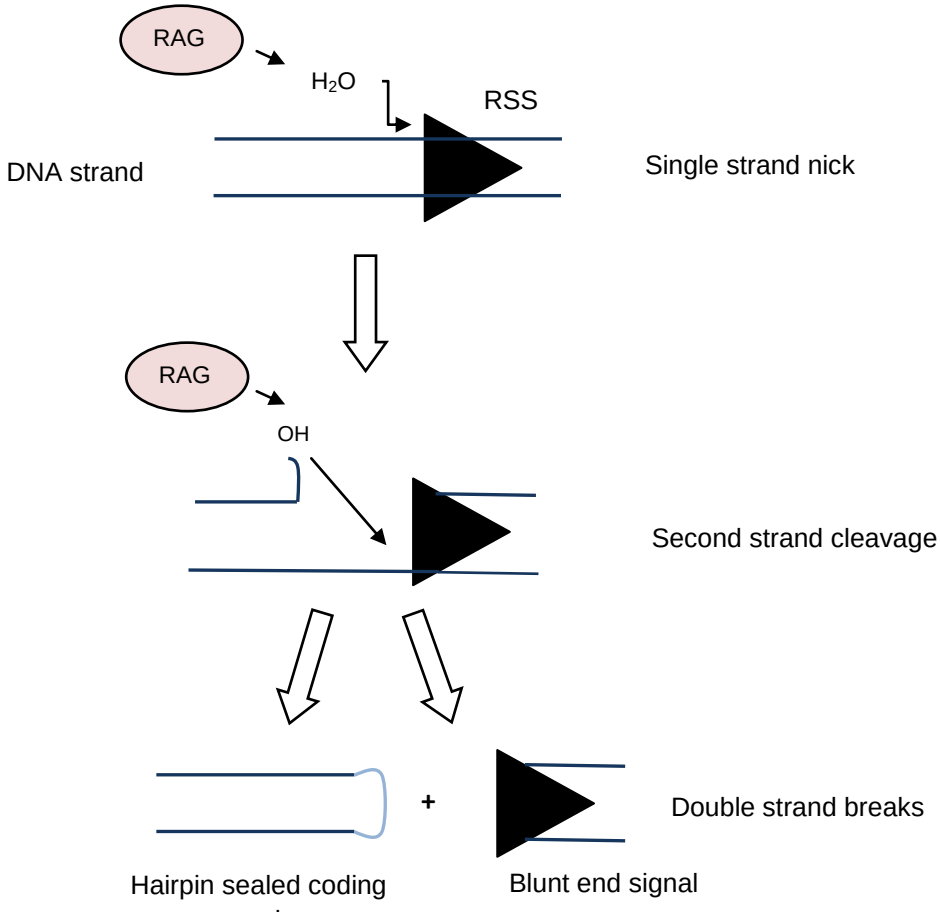


Figure 1-9: Steps involved during cleavage of the DNA by RAG enzymes. The first nick is introduced on one of the DNA strands adjacent to the RSS by hydrolyzing the phosphodiester bond catalyzed by RAG enzymes. As a result of nicking OH^- is generated that attacks on the opposite strand of the DNA resulting in a double stranded break. Later there is a formation of hairpin at the end of both coding and blunt ends of the DNA.

1.3.2.2 Joining phase

The first step to join the broken DNA is the resolution of hairpin loops at the end of coding DNA which is mediated by DNA dependent protein kinase (DNA-PKs) and Artemis (Ma, Pannicke et al. 2002, Rooney, Alt et al. 2005). These proteins are structure specific and process hairpin loops. It results in the processing of coding and signal ends and subsequent joining via classical non homologous end joining (NHEJ) (Taccioli, Rathbun et al. 1993, Ramsden and Gellert 1995). Various important proteins involved in this DNA repair pathway are: Ku70, Ku80, XRCC4, DNA ligase IV and XLF protein (Gellert 2002). At the time of

joining, random non templated nucleotides (N) are added at the junction mediated by Pol X family of polymerases (TdT, polμ and polλ) (Bertocci, De Smet et al. 2006). N nucleotides are added before the ends are joined randomly mediated by the terminal deoxynucleotidyl transferase (TdT) enzyme. The TdT enzyme is maximally expressed at the progenitor (pro) B cell stage therefore adds random nucleotides between both *IGHV-IGHD* and *IGHD-IGHJ* junctions (Desiderio, Yancopoulos et al. 1984). P (palindromic) nucleotides can also be inserted at the junctions. RAG proteins can generate DNA hairpins at the coding ends of V, D and J gene segments after which Artemis enzyme catalyzes a single stranded cleavage at a random point within the coding sequence. As a result, a single stranded tail is formed from a few nucleotides of the coding sequence plus the complimentary nucleotides from the other DNA strand. In most light chain gene rearrangements, DNA repair enzymes then fill the complimentary nucleotides on the single stranded tails, which would leave short palindromic nucleotides at the joint if the ends were rejoined without any further exonuclease activity. Because the total number of nucleotides added by these processes is random, added nucleotides can disrupt the coding sequence beyond the joint. Such frameshifts may result in non productive DNA gene rearrangements (Figure 1-10). At the end of V(D)J gene rearrangement two joints are formed: the coding joint as a result of joining of two gene segments while signal joint represents fusion of two RSSs as shown in Figure 1-11 (Schatz and Ji 2011, Schatz and Swanson 2011).

A: Productive gene rearrangement

DNA ***IGLV*2-14** tgc agc tca tat aca aac agc agc gtg ata ttc ***IGLJ*2**

Protein C S S Y T N S S V I F

(Functional)

B: Non productive gene rearrangement

DNA ***IGLV*2-14** tgc agc tca tat aca aac agc agc **g**gt gat att c ***IGLJ*2**

Protein C S S Y T N S S G D I

(Non Functional)

Figure 1-10: Examples of productive and non productive CDR-3 junctions as a result of gene rearrangement between gene segments *IGLV*2-14 and *IGLJ*2.

(A) CDR-3 junction is defined as productive if the reading frame of triplet codons remains intact and results in a functional protein. **(B)** Random addition or removal of nucleotides (in this example 'g' is inserted) generates non productive gene rearrangement as the reading frame of triplet codon at the junction has been disrupted consequently functional protein cannot be synthesized from this sequence.

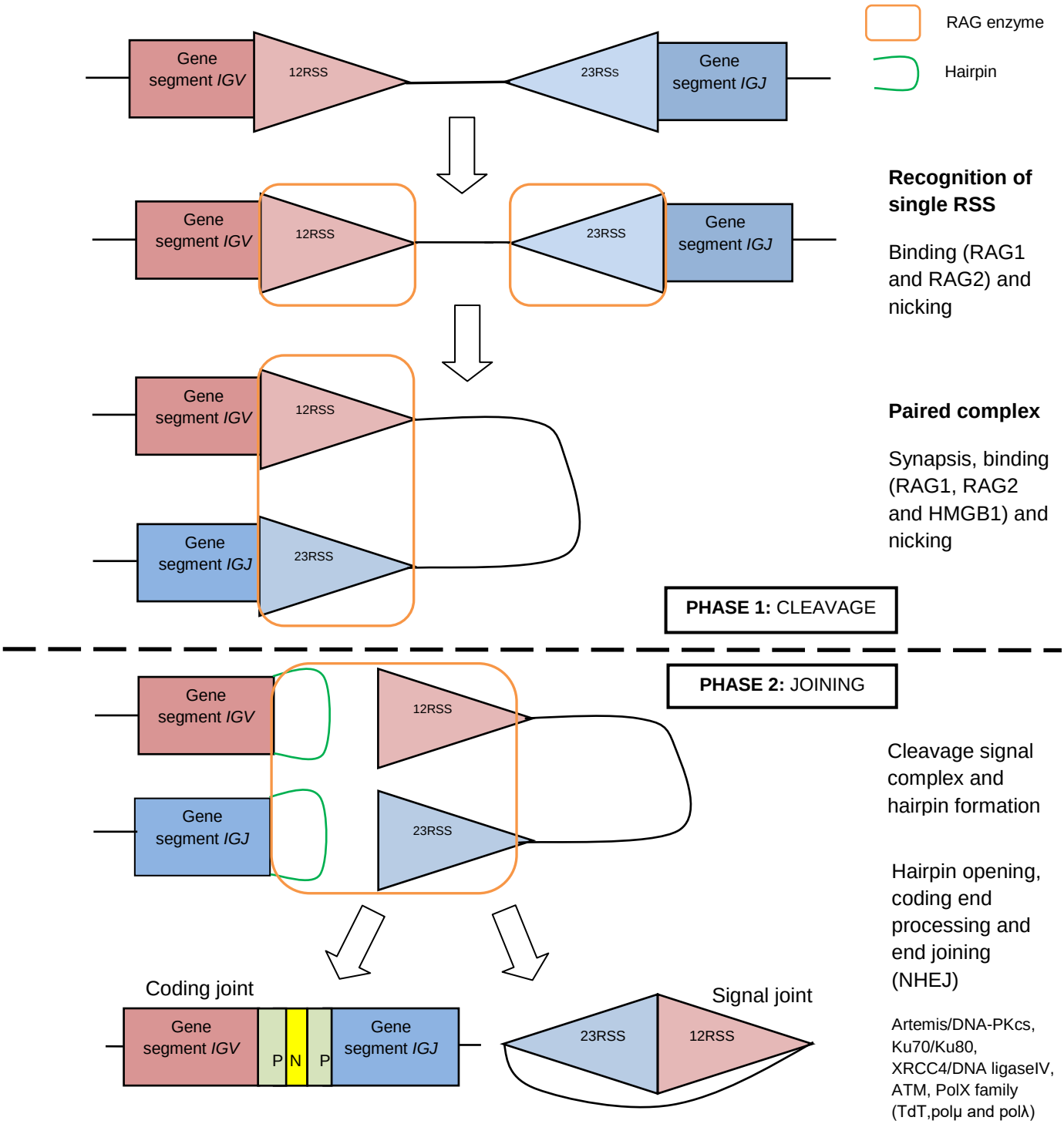


Figure 1-11: Overview of various stages involved during V(D)J gene rearrangement process. Gene segments are flanked by either 12RSS or 23RSS. In the first phase involving cleavage, RAG1 and RAG2 proteins bind to 12RSS or 23RSS and form 12 signal complex or 23 signal complex respectively. Capture of a second RSS (synapsis) result in the formation of the paired complex, within which RAG proteins induce double strand breaks between the gene segments and RSSs. In the second phase that involves joining, the RAG proteins along with other repair proteins rejoin the DNA ends. Gene segments ends typically undergo P and N nucleotides addition by TdT and nucleotides loss by exonucleases before joining to form the coding joint. RSS ends are typically joined without processing.

1.4 B cell development

The development of B cells takes place in the fetal liver and the adult bone marrow both in mice and humans (Ryser and Vassalli 1974, Galy, Travis et al. 1995). Haematopoietic stem cells (HSCs) are pluripotent cells residing in the bone marrow and are precursors of all types of blood cells. Common lymphoid progenitors (CLPs) are descendents of HSCs and progenitors of T, B and natural killer (NK) cells with little or no capacity to differentiate into nonlymphoid lineages such as myeloid and erythroid cells (Galy, Travis et al. 1995). The developmental relationships between HSCs, CLP and potential lymphoid lineages are shown in Figure 1-12.

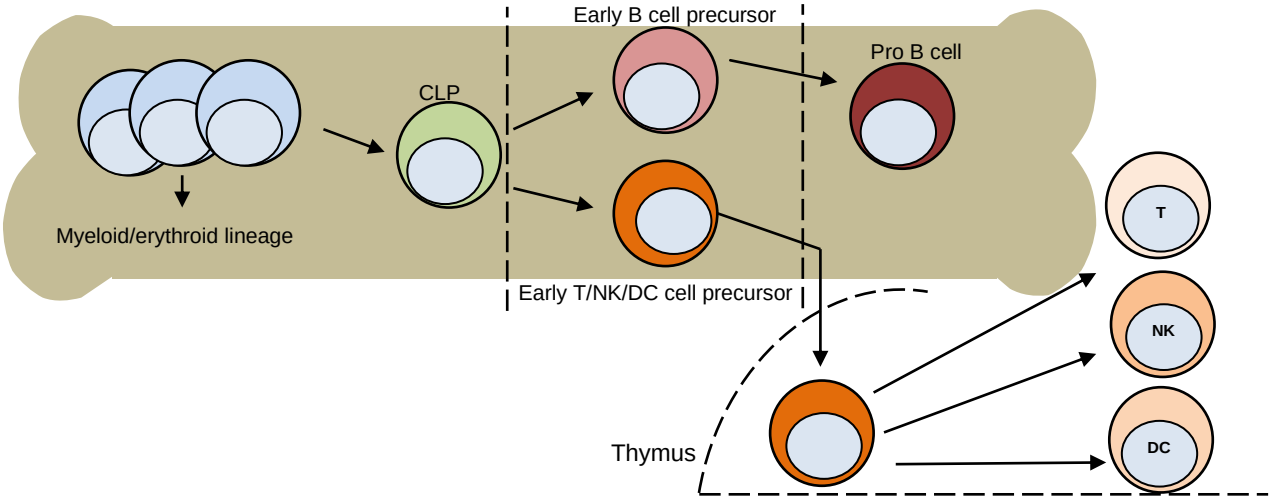
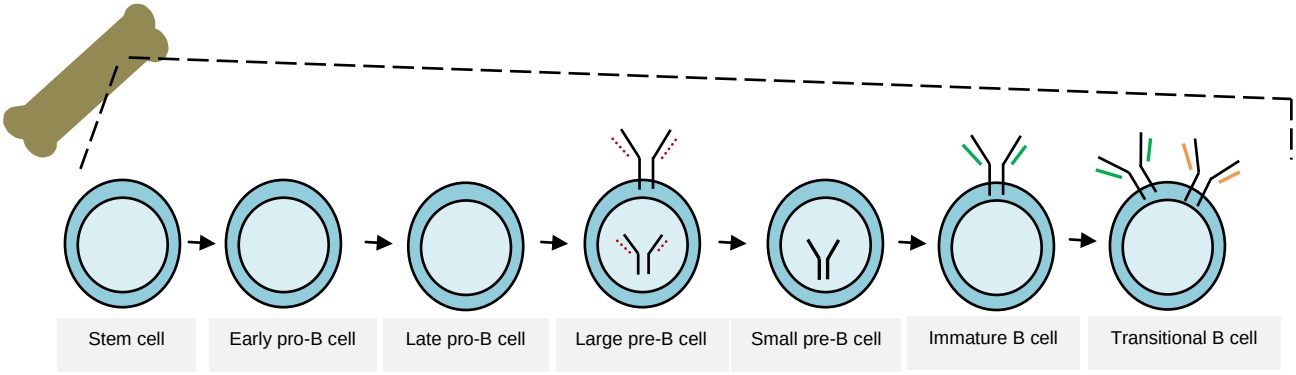


Figure 1-12: Derivation of various blood cells lineages from HSCs. The CLP can differentiate into the specific lineage of both B cells in the bone marrow and a lineage that retains the potential to develop into T, NK and DCs in thymus.

Early B cells are characterized by the initiation of *IGHD-IGHJ* gene rearrangement at the *IGH* locus and the expression of B lineage specific proteins such as VpreB and Igα (Bertrand, Billips et al. 1997, Davi, Faili et al. 1997, Dworzak, Fritsch et al. 1998). The CLP can also differentiate into tri lineage precursor cells which are defined by expression of CD7 (cluster of differentiation 7) on the cell surface. Further maturation and differentiation of

these tri lineage precursors into progenitors of the specific lineages of dendritic (DC), T and NK takes place in the thymus. The process of B cell development takes place within the bone marrow in adult humans (Spits, Blom et al. 1998). The characteristic phenotypic and temporal features of gene rearrangements at distinct developmental phases are summarized in the Table 1-3.

Table 1-3: Various stages of the B cell development starting in the bone marrow



IGH chain genes	Germline	*IGHD-IGHJ* rearranging	*IGHV-IGHDJ* rearranging	*IGH* rearranged	*IGH* rearranged	*IGH* rearranged	*IGH* rearranged
IGK/IGL chain genes	Germline	Germline	Germline	Germline	*IGKV-IGKJ* rearranging	*IGK* rearranged	*IGK* rearranged
Surface Ig	Absent	Absent	Absent	µ chain transiently at surface as a part of pre-BCR	Intracellular µ chain	IgM expressed on cell surface	IgM and IgD made from alternatively spliced H-chain

V(D)J gene rearrangement takes place in a stepwise fashion during B cell development (Alt, Yancopoulos et al. 1984, Ehlich, Schaal et al. 1993, Ehlich, Martin et al. 1994, ten Boekel, Melchers et al. 1995). It begins at the *IGH* locus in early pro B cells when *IGHD* gene segment is rearranged with *IGHJ* gene segment on both alleles as evident by the juxtaposition of gene segments for efficient long distant interaction (Alt, Yancopoulos et al. 1984, Yancopoulos and Alt 1985). In humans most *IGHD* to *IGHJ* joints are potentially useful because *IGHD* gene segments can be translated in all three reading frames without

encountering a stop codon (Benichou, Glanville et al. 2013). In the next step, *IGHV* gene segment rearranges to *IGHD-J* on only one allele and stops further *IGHV* to *IGHD-J* gene rearrangement (Alt, Yancopoulos et al. 1984). However, if the process of rearrangement on the first allele does not result in a functional protein then it is followed by the gene rearrangement on the second *IGH* allele (Vettermann and Schlissel 2010). The productive exon results in the production of an intact cytoplasmic μ heavy chain (μHC) and is necessary for the transition of pro B cells into precursor (pre) B cells (Lutz, Heideman et al. 2011). However, pro B cells that fail to get a productive gene rearrangement after successive gene rearrangements on either allele undergo apoptosis (Lutz, Muller et al. 2006).

The acquisition of cytoplasmic μHC marks the transition to the next developmental stage of B cell differentiation called pre B cells. In pre B cells, μHC is tested for functionality by pairing with the surrogate light chain (Loken, Shah et al. 1987, LeBien, Wormann et al. 1990, Ghia, ten Boekel et al. 1996). The surrogate light chain is composed of two invariant proteins (VpreB and λ5) that form a tight but non covalent pair with μHC and are expressed on the pre B cell surface as a pre BCR along with invariant proteins Igα/Igβ (Melchers, Karasuyama et al. 1993, Keyna, Beck-Engeser et al. 1995, Kline, Hartwell et al. 1998, ten Boekel, Melchers et al. 1998). Biochemical studies reveal that pre BCR is activated by the ligand independent oligomerization (Bankovich, Raunser et al. 2007). The BCR repertoire is largely shaped at this step as it has been found that μHC with the potential to generate autoantibodies undergo negative selection at this step (Keenan, De Riva et al. 2008). More than half of all μHC are potentially unable to pair with surrogate light chain in spite of having a productive gene rearrangement and thus affect IGH repertoire (ten Boekel, Melchers et al. 1998). The production of surrogate light chains is turned off after the cell surface expression of pre BCR (Melchers, ten Boekel et al. 2000). Pre B cell stage marks a selection checkpoint during B cell development. This checkpoint ensures that early B cell development is stringent and cells have not undergone any abnormal transformation. The selection of B cells at the pre B cell checkpoint is regulated by a balance between the two important transcription factors: basic leucine zipper transcription factor 2 (BACH2) and B-cell lymphoma 6 protein (BCL6). The transcription factor BCL6 mediates positive selection of pre B cells and thus allows the cell to cross this checkpoint. On the other hand, transcription

factor BACH2 induces negative selection. (Swaminathan, Duy et al. 2014). Human pre B cells are classified into two subsets based on their proliferative capacities: pre BI cells and pre BII cells. The pre BI cells are large proliferating cells while pre BII cells are small post mitotic cells (Ghia, ten Boekel et al. 1996). Overall, the expression of pre BCR orchestrates survival, proliferative expansion of early pre B cells expressing μHC and the subsequent developmental transition to the late pre B cell stage and leads to the initiation of gene rearrangement at the *IGK* locus (Shapiro, Schlissel et al. 1993, Vettermann and Schlissel 2010).

The process of gene rearrangement at the *IGK* locus starts on any one allele randomly in pre BII cells and during this process the *IGL* locus remains in germline configuration (Beishuizen, Verhoeven et al. 1994). Once pre B cells have successfully joined *IGKV* and *IGKJ* segments to generate a productive *IGKVJ* exon, its functionality is tested by pairing and expressing with the μHC on the cell surface as a BCR (Alt, Yancopoulos et al. 1984). Pre B cells having IGK chain that could not pair with the μHC or encode for an autoreactive receptor may undergo further secondary gene rearrangements termed receptor editing. As a result of secondary gene rearrangements upstream *IGKV* gene segments may rearrange with downstream *IGKJ* gene segments to create a different specificity (Feddersen and Van Ness 1985, Chen, Prak et al. 1997, Pelanda, Schwers et al. 1997). After receptor editing various gene rearrangements can accumulate in the cell on the chromosome and as episomes as shown in theoretical Figure 1-13.

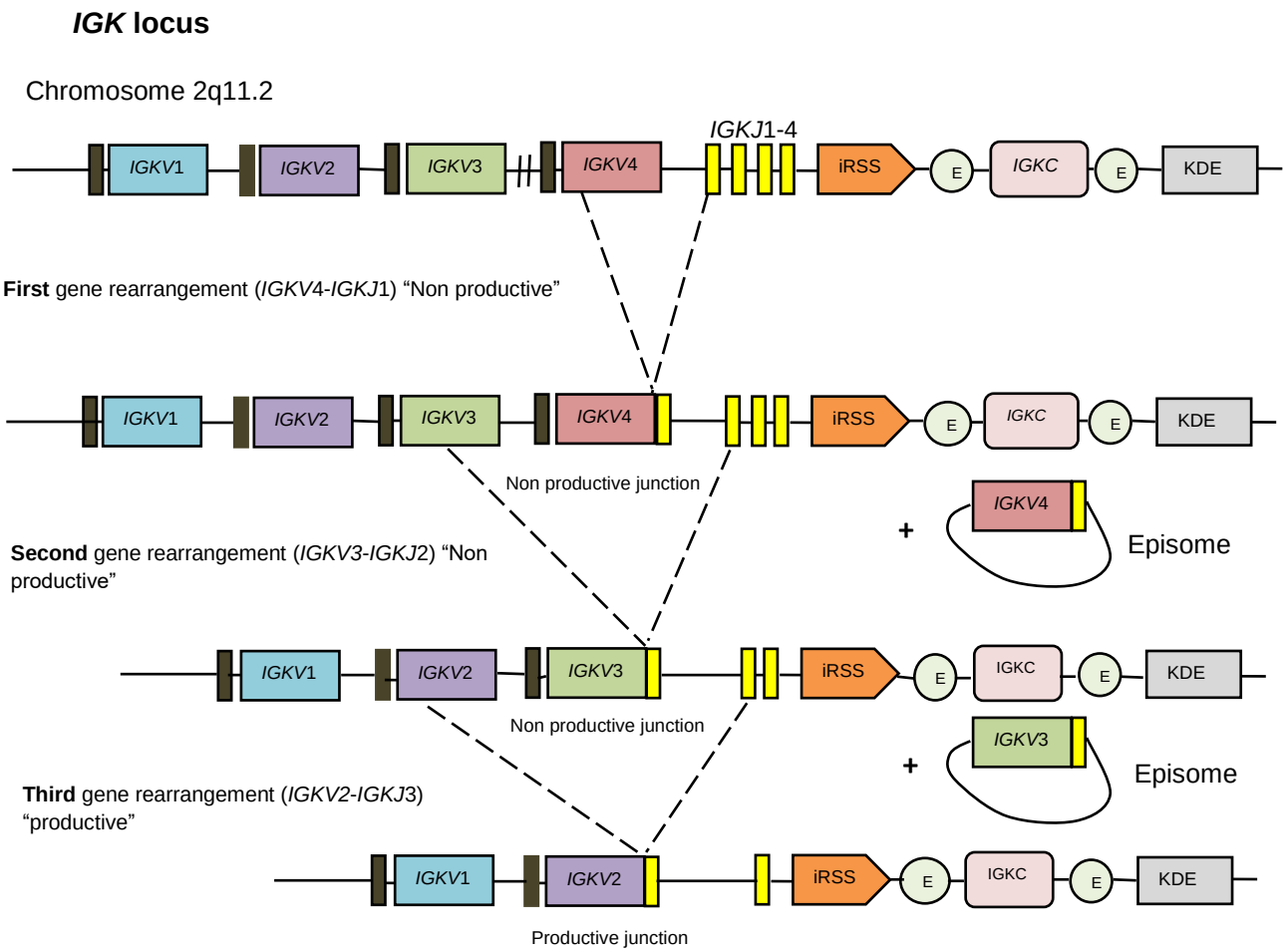


Figure 1-13: Schematic representation of mechanisms leading to the accumulation of multiple gene rearrangements in the same cell. In this example, initial non productive gene rearrangements (involving *IGKV*4 and *IGKV*3) have been followed by the secondary gene rearrangement of an upstream *IGKV2* and a downstream *IGKJ3* gene segment. Theoretically if this gene rearrangement (second) is again either non productive or functionally autoreactive it may lead to the third gene rearrangement on the same allele provided there are unused upstream *IGKJ* gene segments along with the accumulation of excised gene rearrangements.

When either a functional *IGKV-IGKJ* gene rearrangement is obtained or the cell gets exhausted after utilizing all possible options of the secondary gene rearrangements on that allele then it results in the inactivation of the *IGK* locus due to the deletional rearrangement

by kappa deletion element (KDE) as shown in the Figure 1-14. The KDE can be rearranged via one of the two pathways (Siminovitch, Bakhshi et al. 1985, Klobeck and Zachau 1986):

1. KDE undergoes RSS-mediated site-specific recombination with intronic recombination signal sequence (iRSS) located between the distal *IGKJ* and *IGKC* gene segments.
2. The alternative KDE rearrangement takes place to an upstream, unrearranged *IGKV* gene segment.

Both RSS-mediated recombination events result in deletion of the *IGKC* region and its enhancers (iE_K and $3'E_K$); though in the latter instance the *IGKV-IGKJ* joint is also removed. Thus as a result the *IGK* locus loses its potential for further gene rearrangements or expression (Langerak, Nadel et al. 2004, Das, Nikolaidis et al. 2009).

Alternatively, in a minor unconventional pathway, the iRSS can rearrange with the RSS flanking *IGKJ* gene segments. If there is no deletion at this joint then it can undergo further rearrangements either to an upstream RSS element flanking *IGKV* gene segment (*IGKV*RSS-iRSS pseudohybrid joint) or to the downstream RSS of the KDE (*IGKJ*RSS-KDE pseudohybrid joint). This joint, can further recombine in two different ways:

1. Recombination between *IGKV* segment and *IGKJ*RSS-KDE pseudohybrid joint results in *IGKV*-iRSS coupling.
2. Recombination between KDE and *IGKJ*RSS-iRSS joint results in *IGKJ*RSS-KDE coupling.

As long as the iRSS of *IGKV*-iRSS pseudohybrid joint is still intact, further recombination to the RSS of KDE, leading to a *IGKV*-KDE coupling, might occur (Langerak, Nadel et al. 2004).

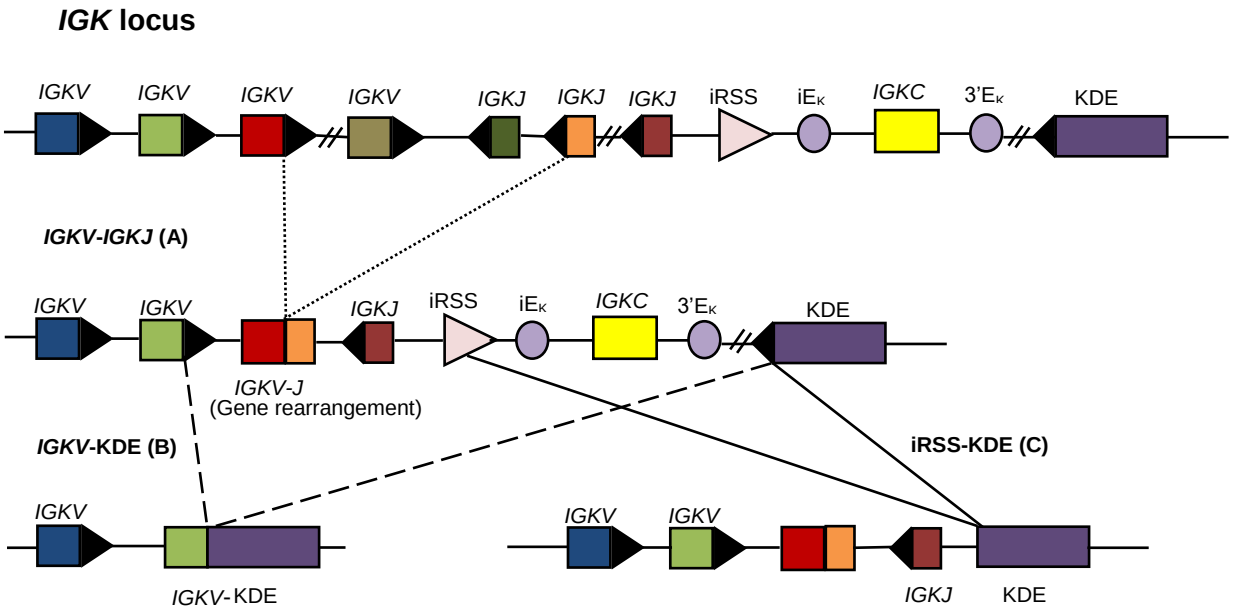


Figure 1-14: Schematic overview of the *IGKV-IGKJ* gene rearrangement and subsequent inactivation of *IGK* locus by the classical pathway. (A) Either functional or exhausted gene rearrangements (*IGKV-IGKJ*) at the *IGK* locus lead to further rearrangements that result in the inactivation of *IGK* locus. **(B)** Recombination of KDE to an isolated heptamer in the intron between the *IGKJ* and *IGKC* segments results in the deletion of the *IGKC* region (solid line), whereas recombination between KDE and upstream *IGKV* gene segment deletes the entire *IGKJ-IGKC* region **(C)**. Both types of KDE rearrangements (B and C) also result in the deletion of *IGK* enhancers (iEk and 3'Ek) and thus preclude further rearrangements in the human *IGK* locus.

Thus inactivation of *IGK* locus helps in establishing allelic and isotypic exclusion (Langerak, Nadel et al. 2004). If a cell fails to get any desired specificity from one *IGK* allele then gene rearrangement starts on the second *IGK* allele. If both alleles fail to produce any functional specificity then gene rearrangement proceeds to the *IGL* locus. Like the *IGK* locus, gene rearrangement can start on the second *IGL* allele if the functional gene rearrangement is not obtained from the first allele (Hieter, Korsmeyer et al. 1981, Korsmeyer, Hieter et al. 1982, van der Burg, Tumkaya et al. 2001). The BCR signal is essential for cell survival therefore failure to rearrange and express both functional IGH with either IGK or IGL results in the death of the cell by apoptosis (Lam, Kuhn et al. 1997).

Immature or transitional B cells express functional IgM BCR on the cell surface. At this stage of B cell development, BCR is first tested for reactivity against self antigens. It has been found that after primary gene rearrangement, up to 10^{11} specificities of BCRs can be generated and some of these can be autoreactive (Grimaldi, Hicks et al. 2005). Cells passing this checkpoint are selected to mature further in the periphery. Cells that react against self antigens undergo apoptosis. However, this early checkpoint was found to be defective in some autoimmune diseases such as systemic lupus erythematousus (SLE) (Su and Rawlings 2002, Yurasov, Wardemann et al. 2005, Pelanda and Torres 2012).

1.5 Molecular factors affecting V(D)J gene rearrangement

Immunoglobulin gene loci harbour multiple gene segments but not all are rearranged with the same frequencies. This suggests there are inherent biases that determine the base level of gene rearrangement irrespective of selection pressure. Following are some factors that could affect the overall process of V(D)J gene recombination.

1.5.1 Differential locus accessibility

Like all cellular DNA, antigen receptor loci exist as nucleosomes. Nucleosomes are the basic building blocks of chromatin. Each nucleosome is composed of an octamer of histone proteins (H2A, H2B, H3 and H4) wrapped by 2 rounds of DNA. Covalent modifications of N-terminal tails of H3 and H4 regulates accessibility to DNA within specific chromatin. Promoters of active genes are enriched with trimethylation of H3K4 (H3K4me3). H3K4me3 is present only at the pro B cell stage. These patterns of H3K4me3 are recognized by the PHD finger of the enzyme RAG-2 (Jiang, Chang et al. 2005). Locus accessibility increases by non-coding RNA transcription. Non-coding RNAs constitute a broad family including miRNAs, siRNAs, piwiRNAs, tiRNAs, cryptic unstable transcripts and long non coding RNAs. These RNAs regulate different aspects of gene regulation. However, in VDJ gene recombination, a role for long non coding RNAs has been implicated. These transcripts are specifically expressed from the particular region of the *IGH* locus that is organized for VDJ recombination (Corcoran 2010). Thus differences in methylation pattern and other epigenetic regulatory processes can vary the accessibility of individual gene segments to the RAG machinery which may contribute to skewed utilization of individual gene segments.

1.5.2 Variation in RSS

Studies have revealed that the efficiency and specificity of V(D)J gene rearrangement process is influenced by nucleotide variability within RSS. The first three base pairs of the heptamer are crucial for RSS binding to RAG proteins (Akamatsu, Tsurushita et al. 1994, Nagawa, Hirose et al. 2004, Schatz and Swanson 2011). The frequency of VDJ recombination can also be changed enormously by variation in the spacer sequence of RSS. It may result in the reduction of binding affinity of RAG proteins and resultant cleavage of

RSS from gene segments is dramatically reduced (Montalbano, Ogwaro et al. 2003). Thus biases in gene usage can also be generated by differences in RSSs associated with individual gene segments.

1.5.3 Transcription factors

Transcription factors are of fundamental importance in regulating various aspects of V(D)J gene rearrangements. The ectopic expression of E2A is sufficient to promote recombinational accessibility of immunoglobulin genes in non B cells (Hsu, Liang et al. 2004, Johnson, Pflugh et al. 2004, Murre 2005). STAT-5 and IL-7 signalling play an essential role in regulating the order of gene rearrangement in pro B cells by suppressing *IGK* gene rearrangement at this stage (Reynaud, Demarco et al. 2008). Ikaros is another important transcription factor that not only affects the expression of RAG enzymes but also regulates the accessibility of various gene segments at the *IGH* locus that can result in biased usage (Reynaud, Demarco et al. 2008). During gene rearrangement at the *IGH* locus, contraction is mediated by transcription factor Pax5 and enzyme PDK1 bringing together *IGHV* with *IGHDJ* (Reynaud, Demarco et al. 2008, Venigalla, McGuire et al. 2013).

1.5.4 Intergenic sense and antisense transcription

VDJ gene recombination is also regulated by intergenic and antisense transcription. Intergenic transcription increases the efficiency of accessibility and thus utilization of segments at the *IGH* locus by recruiting remodelling factors (Chakalova, Debrand et al. 2005). It occurs when the *IGH* locus is actively engaged in VDJ recombination. Intergenic transcripts are expressed on the majority of *IGHD-IGHJ* recombined alleles in pro B cells and is lost following *IGHV* to *IGHD-J* recombination (Johnston, Wood et al. 2006). In addition to this, anti sense transcription also occurs throughout the *IGH* locus before *IGHV* to *IGHD-J* recombination (Chakalova, Debrand et al. 2005). Sense and antisense transcripts are regulated by the Eμ enhancer at the *IGH* locus (Puget, Hirasawa et al. 2015).

1.5.5 Locus contraction and long range interaction

Immunoglobulin loci generally span several hundred kilobases, with distances between *IGHV* and *IGHDJ* segments of up to 500 kilobases. Therefore, locus compaction might improve efficient synapsis of widely separated gene segments (Atchison 2014). Subnuclear

repositioning and locus contraction regulate both recombination and transcription of immunoglobulin loci during lymphocyte development. It has been found that *IGH* and *IGK* loci are preferentially positioned at the nuclear periphery in immunoglobulin non expressing cells such as hematopoietic precursors and T cells. However, in pro B cells as a result of chromosomal contraction these loci are located centrally (Kosak, Skok et al. 2002). The transcription factor Pax5 plays a pivotal role in locus contraction in pro B cells and distal *IGHV-IGHDJ* gene rearrangements (Fuxa, Skok et al. 2004). DNA-binding zinc-finger protein called CCCTC-binding factor (CTCF) is another protein regulating long range chromatin interactions (Ribeiro de Almeida, Stadhouders et al. 2012). In the absence of CTFC, more usage of proximal than the distal gene segments was observed (Ribeiro de Almeida, Stadhouders et al. 2011). YY1 (Yin Yang 1) is another protein regulating not only the assembly of gene segments during VDJ gene recombination but also during class switching by controlling non-coding antisense RNA transcripts, recruitment of proteins to DNA, and interaction with complexes involved in long-distance DNA interactions including the cohesin and condensin complexes (Pan, Papasani et al. 2013, Atchison 2014).

1.5.6 Inhibition of gene rearrangement on the second allele

Every B cell carries 2 alleles (maternal and paternal) for each of *IGH*, *IGK* or *IGL* loci. Therefore, there is a possibility of encoding BCRs of multiple specificities theoretically if allelic exclusion fails. Such polyspecific B cells could recognize a wide range of antigens including self and potentially various pathological conditions could develop. However, it does not normally happen as there are regulatory mechanisms that allow only one *IGH* allele to be expressed with one allele of either *IGK* or *IGL* to constitute a BCR. The surface expression of BCR from a single allele copy of corresponding genetic loci is called allelic exclusion. As a result of allelic exclusion each B cell expresses a BCR of single specificity (Pernis, Chiappino et al. 1965, Vettermann and Schlissel 2010).

In pro B cells, the process of gene rearrangement is initiated on *IGH* locus when *IGHJ* rearrange with *IGHD* on both alleles. However, the next step of *IGHV* gene rearrangement with *IGHD-IGHJ* happens on only one allele. Subsequently the functional expression of μHC on the cell surface by the first allele inhibits *IGHV* to *IGHDJ* gene rearrangement on the

second allele by feedback inhibition (Krangel 2009, Brady, Steinel et al. 2010). Various mechanisms including DNA methylation, asynchronous replication and epigenetic modifications have been implicated in establishing allelic exclusion (Goldmit, Ji et al. 2005). Chromosomal decontraction and centromeric recruitment of the non-functional allele also play a significant role in the maintenance of allelic exclusion because the RAG complex eventually becomes reactivated during development and/or in peripheral lymphocytes (Roldan, Fuxa et al. 2005).

1.6 Censorship of B cells in the bone marrow

As discussed above, specificities of BCRs are generated by imprecise V(D)J joining that may result in non functional or autoreactive receptors. It has been found that approximately 50% to 75% of immature B cells can recognize self antigens both in mice and humans. These self reactive B cells have shown polyreactivity and anti nuclear specificities. However, at later developmental stages such as the transitional and mature naïve B cells there is much lower frequency (approximately 20% to 40%) of autoreactive B cells in the spleen and blood suggesting transitional B cells marks a checkpoint for identifying and removing autoreactive B cells (Grandien, Fucs et al. 1994, Wardemann, Yurasov et al. 2003). Some autoimmune diseases such as SLE and rheumatoid arthritis (RA) are associated with an increase in polyreactive B cells both immature and naïve suggesting some early defects in the mechanism of central tolerance (Samuels, Ng et al. 2005, Yurasov, Wardemann et al. 2005).

On the other hand, some autoreactive B cells having low avidity for self antigens can bypass this early checkpoint (Hayakawa, Asano et al. 2003, Wardemann, Yurasov et al. 2003, Wen, Brill-Dashoff et al. 2005). Such weak self reactive B cells could be considered beneficial potentially as they provide immunity against a broad spectrum of foreign antigens during infection (Mouquet, Scheid et al. 2010).

This central tolerance is a fundamental phenomenon that ensures negative selection of a large proportion of autoreactive B cells produced during initial stages of B cells development. However, how this negative selection is achieved and regulated is still an active area of research. In 1959, Frank Macfarlane Burnet proposed the hypothesis of clonal selection. According to this concept, specificities of B cells are predetermined by random V(D)J gene rearrangement and after recognition of foreign antigens the selected clones are expanded. On the other hand, interaction of B cells with self antigens is detrimental to B cells. Based on this theory it was assumed that immune response is only regulated by either proliferation or deletion of B cells based on the recognition of foreign or self antigens respectively (Ada 2008).

Further studies suggested that proliferation and expansion are not the only two options for regulating tolerance and deciding the future of particular B cells. It was later proposed that the fate of B cells is determined according to the relative affinity of B cells for self antigens. The antigen recognition of ubiquitous self antigens with high avidity by immature B cells can arrest further development of B cells and results in the apoptosis of B cells (Nemazee and Buerki 1989, Hartley, Crosbie et al. 1991, Hartley, Cooke et al. 1993). On the other hand, recognition of self antigens with low avidity does not halt the development process but causes a shortened life span and deletion in the periphery (Goodnow, Crosbie et al. 1988).

It is now considered that the elimination of autoreactive B cells to maintain central tolerance is achieved by three major mechanisms in the bone marrow: receptor editing (Gay, Saunders et al. 1993, Tiegs, Russell et al. 1993, Prak, Trounstine et al. 1994, Retter and Nemazee 1998, Nemazee and Weigert 2000), anergy (Nossal and Pike 1980, Adams, Basten et al. 1990, Fulcher and Basten 1994, Mandik-Nayak, Bui et al. 1997, Roark, Bui et al. 1997) and clonal deletion (Nemazee and Buerki 1989, Norvell, Mandik et al. 1995, Nossal and Pike 2007, Goodnow, Crosbie et al. 2009).

The importance of clonal deletion is well documented for self antigens major histocompatibility complex (MHC) class 1, red blood cells (RBCs) and DNA (Nemazee and Buerki 1989, Okamoto, Murakami et al. 1992, Chen, Nagy et al. 1995). Anergy is the state of non-responsiveness to the antigen. Anergic B cells cannot respond to cellular cognate antigens under optimal conditions of stimulation (Nemazee 2006).

The data accumulated over ensuing years has revealed that clonal deletion and anergy are not the only mechanisms to get rid of autoreactive B cells. Rather clonal deletion is a default process that operates when receptor editing fails to silence the specificity of autoreactive B cells (Pelanda and Torres 2012). The concept of receptor editing was first proposed by the two independent research groups led by Martin Weigert and David Nemazee in 1993 (Gay, Saunders et al. 1993, Tiegs, Russell et al. 1993). Receptor editing means if an autoreactive B cell encounters with self antigens, secondary gene rearrangements can be initiated via signalling through the autoreactive receptor. These secondary gene rearrangements destroy the primary receptor gene rearrangement that confers autoreactivity and replace it with a

new innocuous receptor gene rearrangement. So in this way by receptor editing autoreactivity can be avoided without losing the cell (Nemazee 2006). The secondary gene rearrangements usually involve rearrangement of upstream *IGKV* or *IGLV* gene segments with downstream *IGKJ* or *IGLJ* respectively (Tachibana, Haruta et al. 1999, Monestier and Zouali 2002). Thus secondary gene rearrangements alter the BCR specificity and extinguish autoreactivity, allowing the primary B cell repertoire to develop in the secondary lymphoid organs (Radic and Zouali 1996).

The process of secondary gene rearrangements can occur on either allele or locus. As a result of secondary gene rearrangements allelic and isotypic exclusion can be violated and resulting cells would be polyspecific (Vettermann and Schlissel 2010). Three different outcomes of potential BCRs could be expected theoretically (Figure 1-15 A-C):

Single non autoreactive specificity: If the secondary gene rearrangement takes place on the same *IGK* allele then resulting cells would only have single non autoreactive specificity: innocuous BCR (Figure 1-15A).

Allelically included dual specificities: If the secondary gene rearrangement takes place on the second *IGK* allele (previously non rearranged) then resulting cells would have allelically included dual specificities: autoreactive and innocuous BCRs (Figure 1-15B).

Isotypically included dual specificities: If the secondary gene rearrangement takes place at *IGL* locus then resulting cells would have isotypically included dual specificity: both autoreactive and innocuous BCRs (Figure 1-15C).

Isotypically included (IGK^+IGL^+) dual specificities expressing B cells have also been reported both in healthy humans and mice (Rezanka, Kenny et al. 2005). Dual light chain expression is an indication of B cells going through the process of receptor editing (Pauza, Rehmann et al. 1993, Ghia, Gratwohl et al. 1995, Giachino, Padovan et al. 1995, LeBien 2000, Pelanda 2014). Such B cells expressing dual specificities could be considered as “trojan horses” where expression of the innocuous receptor facilitates maturation, activation and terminal differentiation of B cells and the autoreactive receptor can recognize self antigens and secrete autoantibodies (Pelanda 2014).

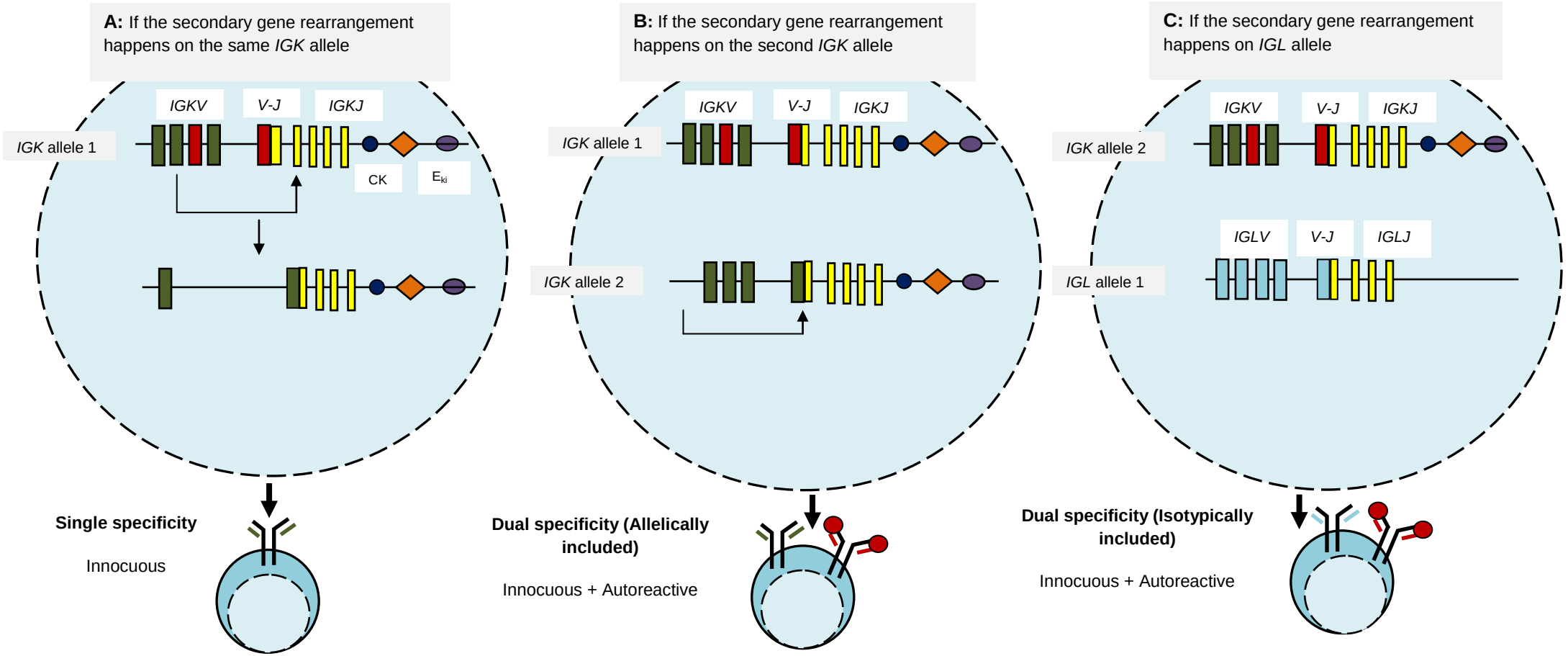


Figure 1-15: Consequences of the secondary gene rearrangements. (A) If the secondary gene rearrangements are initiated on the same allele to replace the potentially autoreactive gene rearrangement then it results in non autoreactive gene rearrangement along with other allele in a germline configuration. **(B)** If the secondary gene rearrangement happens on the second allele then a cell can harbour two specificities of same isotype. **(C)** If the secondary gene rearrangement occurs on the other locus then cell could have two specificities of different isotypes.

1.7 B cell responses in the periphery

Transitional B cells represent the final stage of B cell development in the bone marrow before emigration into the periphery. These cells are defined by phenotype expression of IgM and lack CD27 and ABCB1 transporter protein. Transitional B cells can be divided into two subsets (T1 and T2). T2 subset is more mature than T1 (Sims, Ettinger et al. 2005, Wirths and Lanzavecchia 2005). Recent evidence has shown that the T2 subset of transitional B cells in blood show preferential tropism towards the gut. The migration towards gut is mediated by interaction between gut homing receptor β7 on T2 subset with the endothelial ligand called mucosal vascular addressin cell adhesion molecule-1 (MAdCAM-1) that is expressed by endothelium in gut associated lymphoid tissue (GALT). The T2 subset has higher expression of gut homing receptor β7 than naïve B cells. Moreover, these cells were in activated state as evident by phosphorylation of Bruton's tyrosine kinase (Btk), extracellular signal regulated kinase (Erk) and spleen tyrosine kinase (Syk). Activated transitional B cells in GALT may be the precursors of IgA lineage (Vossenkamper, Blair et al. 2013).

Naïve B lymphocytes migrate to the secondary lymphoid organs such as lymph nodes (LN), Peyer's patches (PP) and spleen. Naïve B cells use L-selectin (CD62L) to roll over specialized capillaries called high endothelial venules (HEVs) present in LN, mesenteric lymph nodes and PPs (Bargatze, Jutila et al. 1995). HEVs also produce certain chemokines such as CCL21 and CCL19 to bind CCR7 on naïve B cells and trigger the activation of integrins LFA-1 and/or α4β7 and lymphocyte arrest in HEVs (Warnock, Campbell et al. 2000).

During unimmunized conditions, the secondary lymphoid organ such as LN is comprised of primary follicles of resting naïve follicular B cells surrounded by T cell rich zones and DCs (Figure 1-16). These B cell follicles are also intersected by follicular dendritic cells having high expression of vascular cell adhesion molecule 1 (VCAM-1), intracellular adhesion molecule 1 (ICAM-1), Fc and complement receptors for the retention of immune complexes (Kitano, Moriyama et al. 2011).

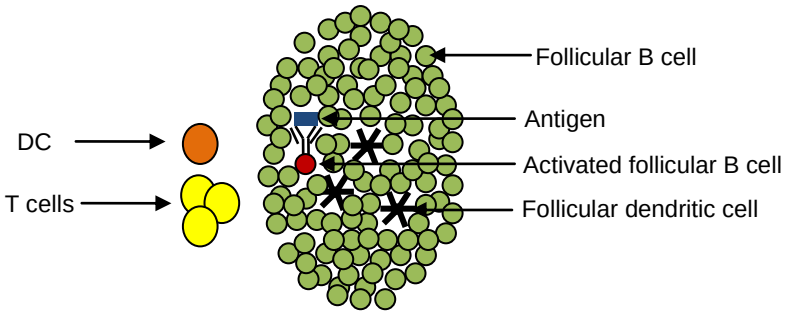


Figure 1-16: Primary B cell follicle in the secondary lymphoid organ. Naïve B cell follicular cells are intersected with FDCs and surrounded by naïve T and DCs.

B cells get activated after exposure to antigens within these primary follicles (Figure 1-16). Later antigen specific activated B cells relocate to the interfollicular zone (between B cell follicles) and this migration is dependent on G protein coupled receptor (GPCR) EB12 (Okada, Miller et al. 2005, Gatto, Paus et al. 2009). On the other side, T cells are primed in the T cell zone after getting activation signals from the DCs. T cells also migrate towards the interfollicular zone by upregulating chemokine receptor CXCR5. These activated B and T cells interact and form stable interactions (CD40 and CD40L) with each other in the interfollicular zone and upregulate transcriptional repressor Bcl6 (Kerfoot, Yaari et al. 2011, Kitano, Moriyama et al. 2011).

The interfollicular zone provides a specialized microenvironment for the differentiation of both T and B cells. A subset of T cells start to differentiate into T follicular helper cells (Tfh) by up regulating CXCR5 and PD-1 (programmed cell death-1) and subsequently Tfh and B cells home into the centre of primary follicle (Kerfoot, Yaari et al. 2011). It has been found if the interaction between B cells and antigen is very strong then in these circumstances, B cells may also differentiate into short lived plasmablasts in the interfollicular zone after getting activation signal from Tfh cells. These plasmablasts provide immediate protection (Paus, Phan et al. 2006).

1.7.1 Establishment of the germinal centre

Antigen specific Tfh cells migrate from the interfollicular region to the centre of the primary follicle on day 3 post immunization, followed by migration of B cells one day later (Kerfoot, Yaari et al. 2011). On day 4 after immunization these antigen specific B cells start to proliferate rapidly within a meshing network of follicular dendritic cells (FDCs). As a result, resident naïve follicular B cells are pushed aside to form an early germinal centre (or secondary follicle), which consists of B cell blasts surrounded by the mantle zone of naïve follicular B cells. On days 5-6, the GC rapidly expands as a result of fast proliferation of B cell blasts. The formation of GC is an imperative phase to establish long lasting humoral immune responses to T cell dependent antigens (Figure 1-17) (Kerfoot, Yaari et al. 2011, De Silva and Klein 2015).

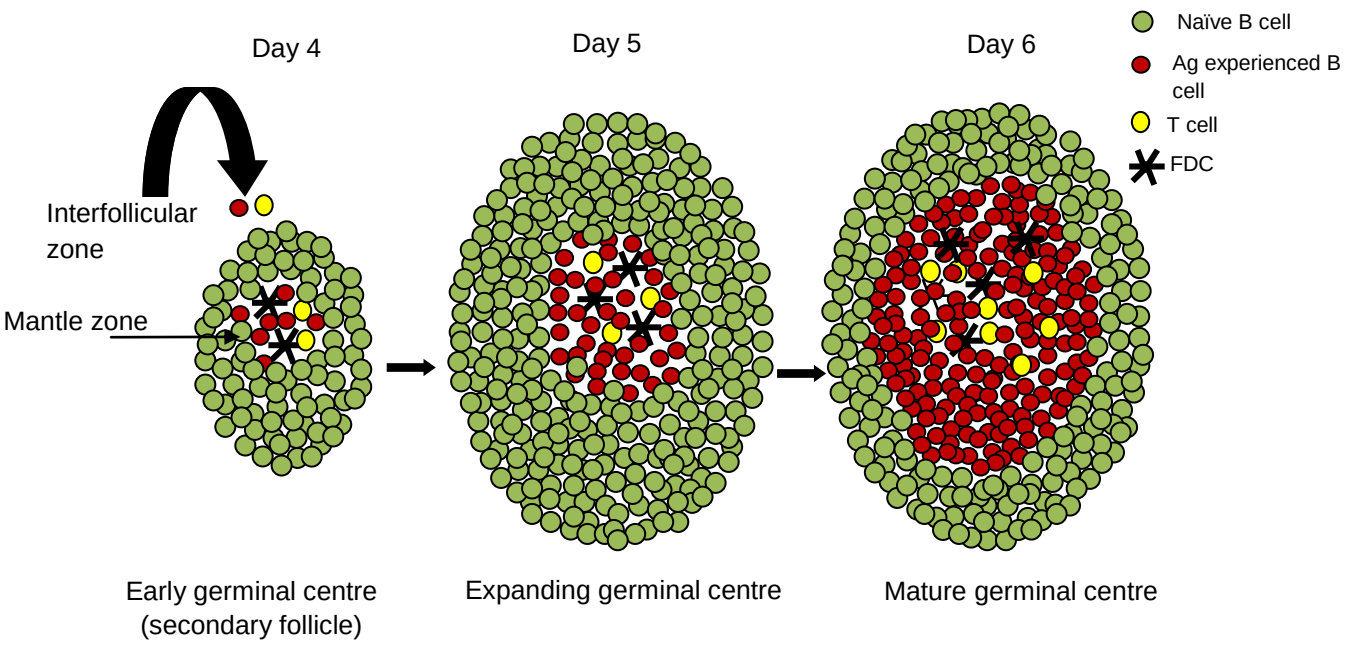


Figure 1-17: Initiation and evolution of the germinal centre. On day 4, antigen specific B cells migrate from the interfollicular region into the centre of the primary follicle, begin to proliferate and push the resident follicular B cells aside to form an early GC, which consists of B cell blasts surrounded by the mantle zone. The structure is also referred to as the secondary follicle. On days 5-6, the GC expands as a result of fast proliferation of antigen specific B cell blasts.

Extensive proliferation takes place until day 7 and B cells in the GC gets polarized into microenvironments: dark zone (DZ) and light zone (LZ) (Victora and Nussenzweig 2012) (Figure 1-18). The DZ is comprised of densely packed blasts called centroblasts that proliferate within a interconnected network of reticular cells that express CXC chemokine ligand 12 (Bannard, Horton et al. 2013). The LZ is sparsely populated by smaller B cells called centrocytes, Tfh, FDCs and macrophages (Victora and Nussenzweig 2012). Centroblasts and centrocytes are polarized into separate zones due to specific chemokines. Centroblasts have higher expression of chemokine receptor CXCR4. The CXCR4 ligand CXCL12 is abundant in the DZ of the GC. On the other hand, centrocytes express chemokine receptor CXCR5 that responds to chemokine CXCL13 produced most abundantly in the LZ (Allen, Ansel et al. 2004). Tfh cells are also attracted to the LZ due to the presence of chemokine receptor CXCR5 on the cell surface which responds to CXCL13 secreted by FDCs (Zotos and Tarlinton 2012).

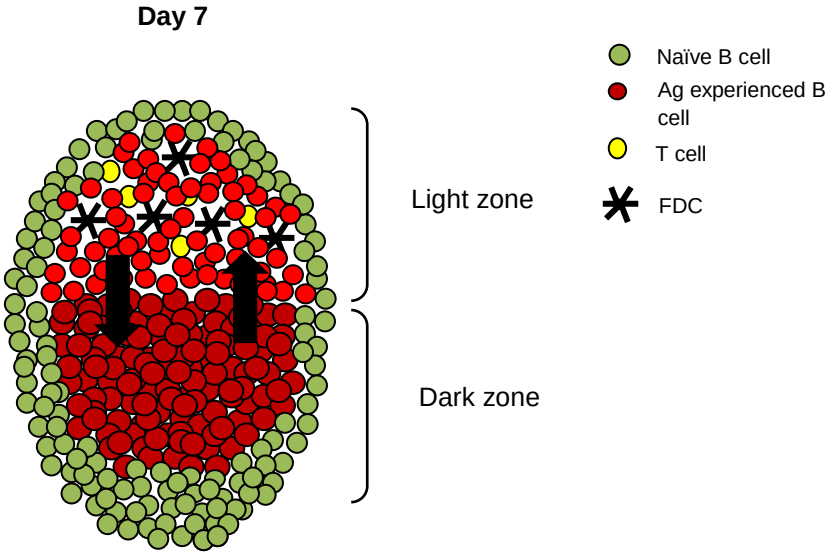


Figure 1-18: Establishment of dark and light zones in the mature germinal centre. On day 7, dark and light zones form which results in the establishment of the mature GC. The dark zone is mainly comprised of densely packed B cell blasts, whereas the light zone contains Tfh, FDCs and blasts. Selected B cells show interzonal back and forth migration.

1.7.2 Selection of high affinity B cells

Centroblasts not only multiply rapidly in the DZ but also undergo hypermutation that introduces point mutations into the variable region of immunoglobulin *IGV* genes and associated amino acid substitutions. This somatic hypermutation (SHM) results from the deamination of deoxycytidines residues to deoxyuridines in *IGV* genes of single stranded DNA (ssDNA) and is mediated by activation induced deaminase (AID) enzyme (Muramatsu, Kinoshita et al. 2000). The SHM process introduces a single nucleotide exchange per 1,000 bases into the rearranged *IGV* gene per division within the DZ of the GC. Thus as a result of SHM a wide range of BCRs is generated having different affinities for antigens (Di Noia and Neuberger 2007, Victora and Nussenzweig 2012).

In the next phase, high affinity antibody producing cells are selected for clonal expansion by Tfh cells in the LZ of GC. B cells capture antigen from FDCs via their BCR and after processing present to Tfh cells via MHC class II on the cell surface. High affinity BCRs capture more antigen that results in higher density of peptide presentation on MHC class II and thus drives positive selection (Victora, Schwickert et al. 2010, Gitlin, Shulman et al. 2014). During interaction, influx of Ca^{++} increases in Tfh cells along with co expression of cytokines IL-4 and 21 (Shulman, Gitlin et al. 2014). B cells in the LZ that are unable to retrieve sufficient antigen undergo apoptosis. After positive selection, a subset of B cells recirculates to the DZ or may undergo class switch recombination (CSR) before migrating to the DZ. In the DZ, these selected cells proliferate again and undergo further SHM. The extent of back and forth migration between the DZ and LZ is directly proportional to the affinity of B cells for antigen and the amount of the antigen they present to Tfh cells. Eventually a pool of high affinity memory B cells and plasmablasts is generated in the LZ (Allen, Okada et al. 2007, Gitlin, Shulman et al. 2014). CSR will be discussed in the section 1-8 in the context of IgA. Various events of the GC reaction are summarized in Figure 1-19.

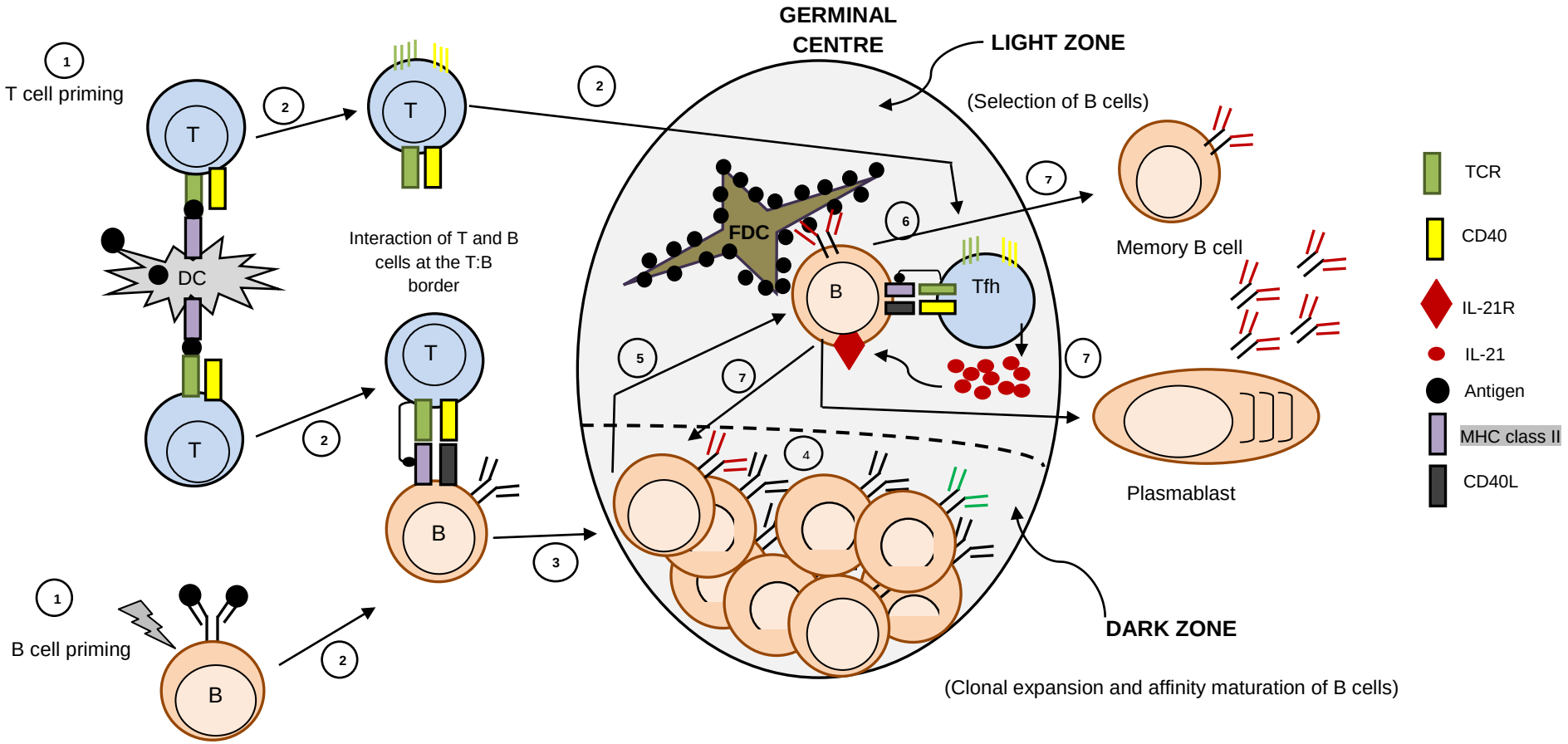


Figure 1-19: Schematic representation of dynamics and selection in the germinal centre reaction at various stages. (1) Sensitization of T and B cells in the respective T and B follicles **(2)** Interaction of sensitized B and T cells at the T:B border **(3)** Entry of B cells into the DZ **(4)** Proliferation of centroblasts **(5)** Entry of centrocytes in the LZ **(6)** Selection of centrocytes **(7)** Differentiation into plasma and memory B cells or migrate back to the DZ

1.8 Immunoglobulin A

The gut mucosal surface is constantly exposed to a wide range of microbiota. It gets colonized with microorganisms shortly after birth (Macpherson 2006). The frequency of bacterial flora increases along the gastrointestinal tract. The proximal parts including stomach, duodenum and jejunum have relatively lower density of the bacteria approximately 10^3 to 10^5 per gram. The distal parts including ileum or colon may contain up to 10^{10} to10^{12} microorganisms per gram. These microbes constitute highly diversified communities belonging to more than 1000 species. This indigenous bacterial flora is mainly comprised of obligate anaerobes (*Bacteriodes* and various species of bifidobacteria, fusobacteria, peptostreptococci) as well as both obligate and facultative aerobes, such as enterobacteria and lactobacilli. These commensals compete with pathogens for the attachment to the gut and also stimulate the growth of intestinal epithelial cells and thus facilitate the development of a healthy immune system (Macpherson and Harris 2004, Rakoff-Nahoum, Paglino et al. 2004, Rhee, Sethupathi et al. 2004, Hooper, Littman et al. 2012, Kato, Kawamoto et al. 2014). It has been reported recently that IgA coating identifies inflammatory commensals that preferentially drive intestinal disease. This property could be advantageous in translational research as by specifically designing antibodies against such bacteria progression of colitis or even initiation might be checked (Palm, de Zoete et al. 2014). Overgrowth of segmented filamentous bacteria has been reported throughout the small intestine in IgA deficient mice (Suzuki, Meek et al. 2004). IgA can also cause activation of eosinophils and subsequent degranulation and release of reactive oxygen species (ROS) suggesting implications in allergy (Pleass, Lang et al. 2007).

The gastrointestinal tract harbours distinct mucosal lymphoid microenvironments: organized mucosa associated lymphoid tissue (MALT), diffuse connective tissue of lamina propria (LP), regional mesenteric lymph nodes, intraepithelial lymphocytes (IELs) (Brandtzaeg and Pabst 2004, Brandtzaeg 2009, Spencer, Klavinskis et al. 2012). MALT can be defined by the juxtaposition of organized lymphoid tissue to specialized lymphoepithelium called follicle associated epithelium (FAE). The FAE acts as a barrier between the intestinal lumen and the internal environment of the body. Most MALT is located in the gastrointestinal tract called

GALT in either clusters of PP or isolated lymphoid follicles (ILFs) (Gibbons and Spencer 2011).The FAE is interrupted by some specialized epithelial cells called microfold cells (M cells) that are involved in sampling antigens from the lumen and presenting to lymphocytes in the GALT (Owen 1977, Kadaoui and Corthesy 2007). The GALT can also be induced in response to an infection for example MALT is acquired in the stomach in response to *Helicobacter pylori* infection (Wotherspoon, Ortiz-Hidalgo et al. 1991, Acheson and Luccioli 2004). On the other hand, the LP is composed of diffused connective tissue and is infiltrated with effector cells that includes cytokine producing T cells, immunoblasts and plasma cells (Spencer, Klavinskis et al. 2012).

During a mucosal immune response IgA is the predominant isotype produced (Elson and Ealding 1984). The distinct protein fraction of IgA was first noticed in 1958 when Gugler and co workers reported that not all myeloma proteins can be classified according to the prevailing classification system that is based on the carbohydrate contents: the 7S (low carbohydrate content) and the 16S (high carbohydrate content). They found in milk secretions there were some proteins that had relatively higher carbohydrate contents (Gugler, Bokelmann et al. 1958). A year later Heremans and co-workers found there is a protein in secretions having a different characteristic mobility pattern. This fraction of protein tends to migrate towards β globin region during immunoelectrophoresis. Subsequently, this fraction of protein was defined as IgA (Heremans, Heremans et al. 1959). Later studies revealed that the presence of this isotype is predominant in exocrine secretions (Tomasi, Tan et al. 1965, Hanson and Brandtzaeg 1993). Although in the serum, IgA is a relatively small component of antibodies but it is found in much abundance in normal mucosal surfaces where it may account for at least 70% of all immunoglobulins produced in mammals (Macpherson, McCoy et al. 2008).

IgA is often regarded as a non inflammatory and neutralizing antibody. It down modulates the expression of proinflammatory bacterial epitopes. It also blocks the attachment of microbial components to the cell surface, facilitates intraepithelial neutralization of incoming pathogens and microbial inflammatory products (Fagarasan, Kinoshita et al. 2001, Phalipon, Cardona et al. 2002, Peterson, McNulty et al. 2007). It also helps in eliminating the

pathogenic microorganisms that breach the epithelial barrier by transporting them back into the lumen through polymeric Ig receptor (pIgR). It also promotes clearance of pathogens by antibody dependent cell mediated cytotoxicity (ADCC) via FcαRI receptor. FcαRI is a Fc receptor for IgA that is expressed by DCs, neutrophils and other phagocytes (Geissmann, Launay et al. 2001, Pasquier, Launay et al. 2005).

1.8.1 Mucosal and systemic IgA

Systemic IgA generally exists in the monomeric form though IgA polymers also exist in serum in humans. In contrast, IgA is present in mucosal secretions in dimeric form. After production of dimeric IgA by plasma cells it is translocated across epithelial surfaces via interaction between J chain and pIgR protein receptor (Vaerman, Langendries et al. 1998). pIgR is a transporter present on the basolateral surface of epithelial cells mainly present in the intestinal crypts as membrane secretory component. IgA is transported into the lumen with bound secretory component that is produced by apical cleavage of pIgR. J chain of polymeric immunoglobulins enables binding to pIgR (Brandtzaeg 2013). Certain amino acid motifs in Cα2 and Cα3 domains are of considerable importance for pIgR binding and transcytosis of IgA. It was found that disruption of Cys311 in Cα2 resulted in reduced binding of IgA to pIgR. Combined mutagenesis of Phe411, Val413, and Thr414 residues in Cα3 domain obviated interaction of IgA with pIgR (Lewis, Pleass et al. 2005). The receptors for IgA are present on granulocytes, monocytes, macrophages, DCs, eosinophils, FDCs, hepatocytes, hepatic kupffer cells and renal mesangial cells. The receptor FcαRI has been implicated in facilitating internalization of IgA opsonised bacteria by phagocytic cells (Cerutti 2008, Macpherson, McCoy et al. 2008, Brandtzaeg 2013).

1.8.2 IgA subclasses

In humans there are two subclasses of IgA: IgA1 and IgA2 encoded by two functional *IGHC*α1 and *IGHC*α2 gene loci respectively on chromosome 14 (Flanagan and Rabbitts 1982). IgA1 subclass is predominant in serum in monomeric form while IgA2 forms stable dimers and is present in secretions increasing in relative amount from small intestine towards large intestine (Crago, Kutteh et al. 1984, Kett, Brandtzaeg et al. 1986). There are also some structural differences between the subclasses such as a deletion of 13 amino

acids in the hinge region of subclass IgA2 that makes it relatively more resistant to the enzymatic digestion by proteases (Kilian, Reinholdt et al. 1996, Brandtzaeg and Johansen 2005, Macpherson, McCoy et al. 2008). The subclass IgA2 has better Fcα mediated mannose dependent agglutinating properties against enteric microorganisms and is more glycosylated than IgA1 (Tomana, Niedermeier et al. 1976, Woof and Mestecky 2005, Cerutti 2008).

1.8.3 Inductive sites for class switch recombination to IgA

The mucosal immune system can be divided into two compartments based on involvement in immune responses: inductive sites and effector sites. Inductive sites denote the region where naïve B and T cell encounter with antigen and get activated e.g. palatine tonsils, adenoids, PP (located mainly in the distal ileum), MLN and ILFs most abundantly in the large bowel (Brandtzaeg and Pabst 2004, Brandtzaeg 2010). PPs are the major sites of IgA induction followed by the MLN (Macpherson and Slack 2007). The effector sites mark the region where final differentiation to plasma cells takes place e.g. subepithelial lamina propria of various mucosae, stroma of exocrine glands, and the surface epithelia (Brandtzaeg and Johansen 2005, Brandtzaeg 2009, Brandtzaeg 2010).

1.8.4 Mechanism of class switch recombination to IgA

CSR to IgA takes place by the activation of the AID enzyme in activated B cells (Muramatsu, Kinoshita et al. 2000). The immunoglobulin *IGH* locus of mature B cells contains a VDJ exon encoding the antigen binding domain of immunoglobulin. B cells produce intact IgM and IgD receptors through a transcriptional process driven by a promoter upstream of the VDJ exon following rearrangement of the light chain. Therefore, in order to produce downstream isotypes such as IgG, IgA or IgE having identical antigen specificities only constant regions are changed by the process of CSR. The CSR from IgM to IgA takes place as a result of transcription from the intronic α promoter 5' of the *IGH*α locus mediated by AID in the GC. The intervening DNA is excised as a circular DNA and remaining exons splice together to form a functional transcription unit for IgA production as shown in Figure 1-20 (Chaudhuri and Alt 2004, Kumar, DiMenna et al. 2014). The CSR involves an exchange of upstream donor Cμ and Cδ genes with a downstream acceptor C_H gene through a recombinational

process that is guided by switch (S) regions (Kotnis, Du et al. 2009). The S region (1 to 12kb) consists of highly repetitive sequences that are rich in G. Each S region is preceded by a short intronic (I) exon and a promoter that initiates germline transcription on the reception of activation signals (Chaudhuri and Alt 2004, Muramatsu, Nagaoka et al. 2007). During germline transcription, Iα-Cα transcript is produced as result of splicing. The primary sterile transcript physically associates with the template strand of the DNA to form a stable DNA-RNA hybrid (Reaban and Griffin 1990). Subsequently AID deaminates cytosine residues on both strands of S region DNA, thereby generating multiple DNA lesions that are ultimately processed into ds DNA breaks. Fusion of ds DNA breaks at Sα and Sμ through NHEJ pathway induces looping out and deletion of the intervening DNA thereby juxtaposing V_HDJ_H to Cα. This process yields a chromosomal V_HDJ_H-Cα sequence, which encodes the IgA protein, and an extrachromosomal switch circle (Kinoshita, Harigai et al. 2001, Chaudhuri and Alt 2004, Cerutti 2008, Stavnezer, Guikema et al. 2008).

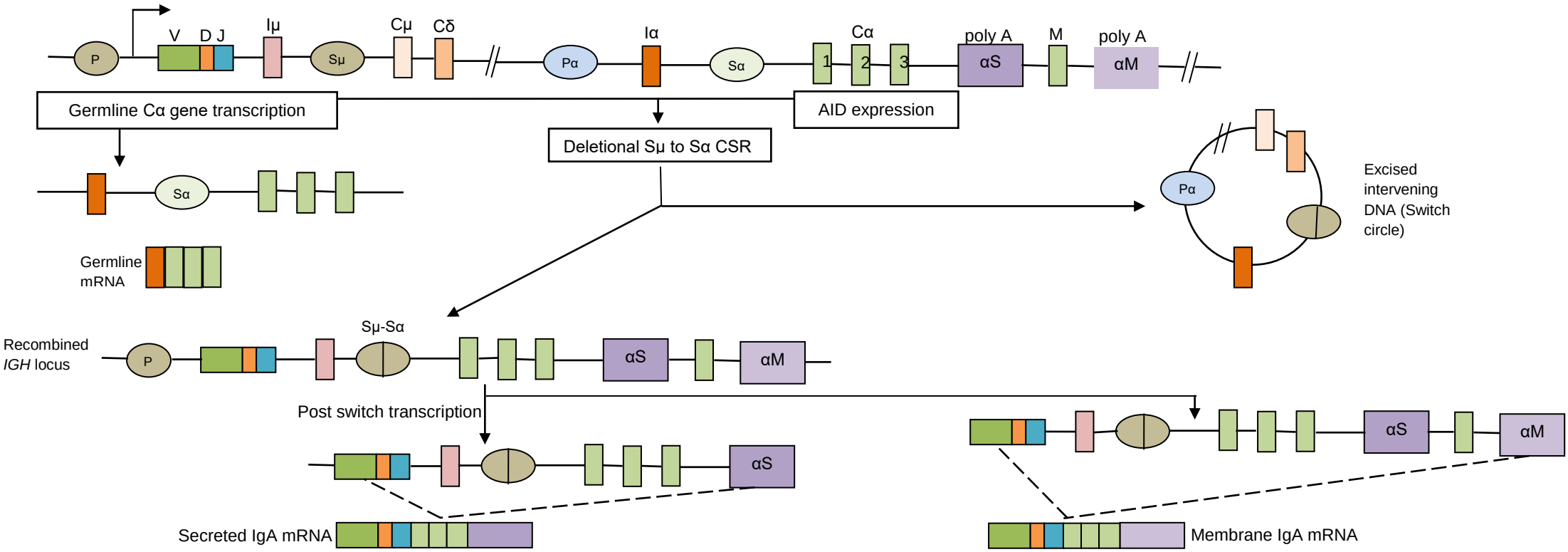


Figure 1-20: Recombinational and transcriptional events during IgA class switching. Appropriate stimuli induce germline transcription of the Cα gene from the promoter Pα of the intronic Iα exon through the switch Sα between Iα and Cα exons. In addition to yielding sterile transcript, germline transcript renders the Cα gene substrate for AID. By generating and repairing DNA breaks at Sμ and Sα, the CSR machinery rearranges *IGH* locus, thereby yielding excision switch circle. Post switch recombination generates mRNAs for both membrane and secreted forms.

1.8.5 T cell dependent and T cell independent IgA

The prerequisite to initiate the process of CSR from IgM to IgA are as follows (Stavnezer 1996, Muramatsu, Kinoshita et al. 2000):

1. Enzymatic help by AID
2. Two activation signals

The first signal is usually provided by the cytokines that start the transcription of the promoter upstream of the *IGHC* exons and the second signal is given by T cells when CD40 on B cells interacts with CD40L on T cells. The important cytokines for IgA class switching are transforming growth factor β (TGF-β), A proliferation inducing ligand (APRIL), IL-2, IL-4, IL-5, IL-6 and IL-10 (Islam, Nilsson et al. 1991, Shockett and Stavnezer 1991, Borsutzky, Cazac et al. 2004). Although the activation signal by T cells is important but CSR can also take place without the T cell help. In such cases, a co stimulatory signal is given by other cells such as macrophages, monocytes and dendritic cells (Macpherson, Hunziker et al. 2001, Barone, Patel et al. 2009, Bergqvist, Stensson et al. 2010, Berkowska, Driessen et al. 2011). On this basis there are two major pathways of IgA production:

1.8.5.1 T cell dependent IgA production

T cell dependent antibody response involves activation of B cells by antigens in the organized lymphoid tissue of PPs, MLNs and ILFs (Fagarasan and Honjo 2003, Macpherson, McCoy et al. 2008). T dependent (TD) IgA generation takes place in the GC when activated B cells get a co stimulatory signal by T cells (Casola and Rajewsky 2006). In humans IgA expressing B cells are mostly generated in the GC of PPs (Lin, Du et al. 2014). The TD IgA is characterized by phenotypic expression of CD27 on the surface of IgA expressing B cells. The GC B cells in PPs are biased for CSR to IgA possibly because of cytokines that drives IgA class switching (Cerutti and Rescigno 2008, Suzuki and Fagarasan 2009). TGFβ is an essential cytokine for IgA class switching. TGFβ along with IL-21 produced by Tfh synergistically enhances both proliferation and differentiation of precursors for IgA plasma cells. However, IL-4 produced by activated T cells suppresses IgA production (Dullaers, Li et al. 2009, Tsuji, Komatsu et al. 2009). The essential steps for the production of IgA dependent on T cells are summarized in Figure 1-21. DCs located beneath M cell

uptake luminal antigens and sensitize Tfh cells in the T cell zones. B cells also process and present invading antigens. The activated B cells in GALT are characterized by the phosphorylation of Btk and Erk, down regulation of CD22, upregulation of signalling proteins (Fos and Jun), and translocation of Lyn to the cell membrane. These sensitized B cells later interact with T cells and get activated after reception of two signals: Specificity signal (MHC class II on B cells interact with TCR) and co stimulatory signal (CD40L on B cells interact with CD40 on T cells). As a result of these interactions, under the influence of TGFβ activated B cells class switch to IgA expressing B cells and this enzymatic process is mediated by AID (Suzuki and Fagarasan 2009, Barone, Vossenkamper et al. 2011).

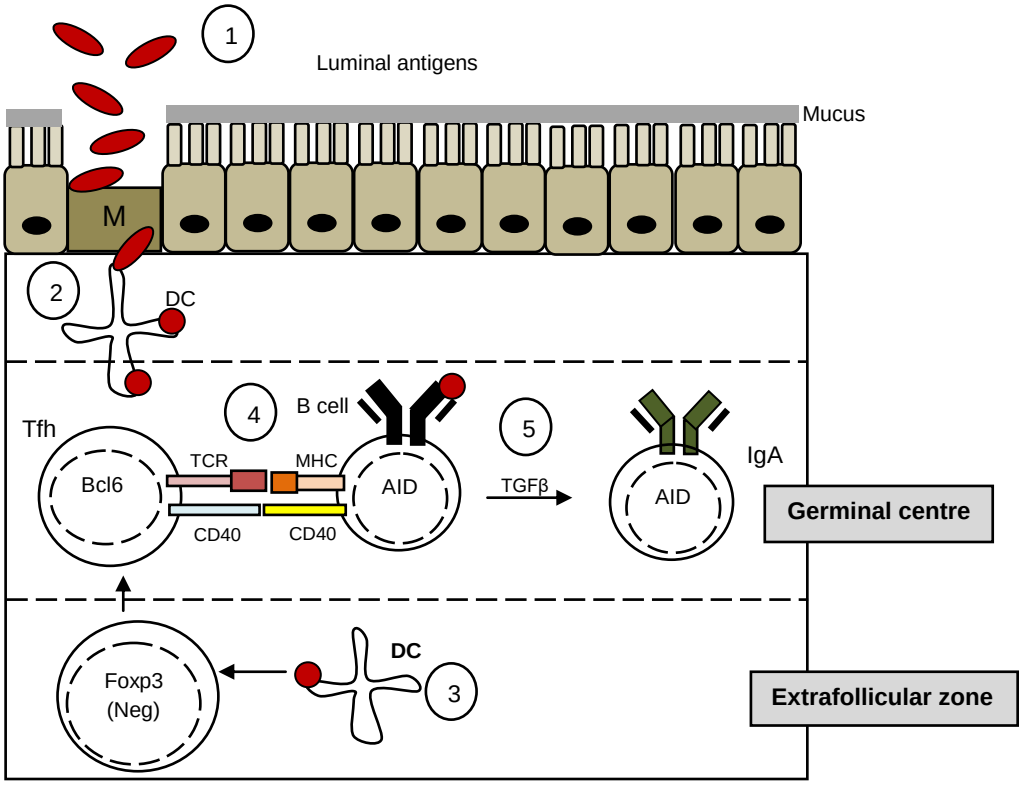


Figure 1-21: Schematic representation of mechanisms for generating T dependent IgA. DCs capture luminal antigens by interacting with M cells (**1** and **2**). Activated DCs present antigens to $CD4^+$ T cells and cause differentiation into Tfh cells that secretes IgA inducing cytokines (**3**). As a result of interaction between activated B and T cells CSR to IgA takes place mediated by AID (**4** and **5**)

1.8.5.2 T cell independent IgA production

It has been suggested that T independent (TID) IgA could be characterized by the absence of phenotypic expression of CD27 on the cell surface. It can be generated in CD40L deficient patients having no T cells for interaction with B cells (Berkowska, Driessen et al. 2011). Production of TID IgA has also been found intact in HIV patients having low CD4 count or patients with severe T cell depletion (He, Xu et al. 2007, Levesque, Moody et al. 2009).

For the production of TID IgA, the second activation signal to naïve B cells is usually provided by members of tumour necrosis factor (TNF) family: B cell activating factor (BAFF) and APRIL along with other cytokines e.g. IL-4 (Litinskiy, Nardelli et al. 2002, Hardenberg, Planelles et al. 2007, He, Xu et al. 2007, Barone, Patel et al. 2009). Various signals required for activation and CSR towards IgA without any involvement from T cells are summarized in Figure 1-22.

DCs and monocytes produce BAFF while APRIL is expressed by monocytes, macrophages, DCs and activated T cells (Mackay, Schneider et al. 2003). B cells express three types of receptors that could respond to BAFF and APRIL: BAFF receptor (BAFF-R), B cell maturation antigen (BCMA) and transmembrane activator (TACI). These receptors have overlapping specificities for APRIL and BAFF. APRIL binds to BCMA and TACI while BAFF interacts with all three receptors (Macpherson, McCoy et al. 2008).

Mutations in the gene encoding TACI have been reported to be associated with certain immunodeficiencies such as common variable immunodeficiency (CVID) and IgA deficiency (Castigli, Wilson et al. 2005, Salzer, Chapel et al. 2005). It has been found in animals defective for APRIL or its receptors have abnormalities related to the plasma cell survival and IgA production (Castigli, Scott et al. 2004, Bossen and Schneider 2006). The importance of APRIL in IgA regulation is also highlighted in studies suggesting APRIL can promote CSR to IgA by inducing the expression of AID that is independent of interaction of CD40 and CD40L (Litinskiy, Nardelli et al. 2002). Thus, it is an alternate pathway of IgA class switching where the help provided by T cells has been substituted by APRIL. Some reports suggest that TID IgA generation takes place at the extrafollicular site in the LP (He,

Xu et al. 2007). However, LP cannot be considered as a convincing inductive site because there is no consensus on the expression of AID in the LP considering AID is the main prerequisite for the initiation of CSR (Shikina, Hiroi et al. 2004, Bergqvist, Gardby et al. 2006, He, Xu et al. 2007). On the other hand, co expression of APRIL and AID in the human GALT suggest GALT could support the production of IgA in TD way but also independently of T cell help (Barone, Patel et al. 2009). Another study conducted on mice supports TID IgA generation occurring at extrafollicular sites in GALT but not in the LP (Bergqvist, Gardby et al. 2006). Another experiment conducted on $CD40^{-/-}$ mice revealed normal levels of IgA plasma cells in the LP although there were no GCs and CSR was evident only in PP (Bergqvist, Stensson et al. 2010).

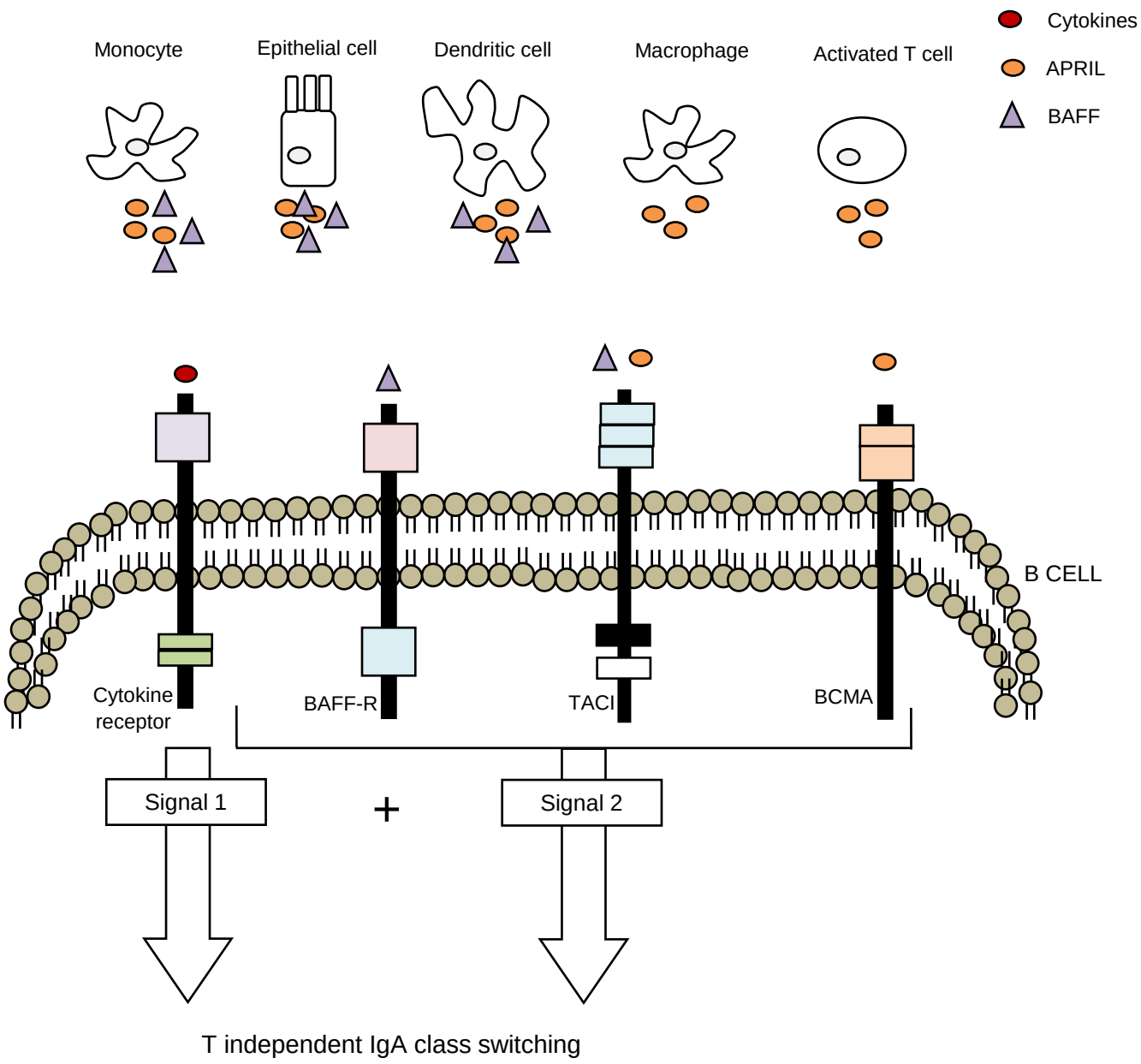


Figure 1-22: Stimulatory signals for CSR to T independent IgA. BAFF and APRIL are expressed by DCs, monocytes and human colonic epithelial cells (following TLR signaling), whereas APRIL can also be secreted by macrophages, and activated T cells. These proteins bind to their receptors, BAFF-R, TACI or BCMA which are all expressed by B cells.

Upon leaving the GALT, B cells return to the blood where they must express the gut homing receptors α4β7, CCR9 and CCR10 in order to emigrate into the small bowel (Hart, Ng et al. 2010). It has been shown that retinoic acid (RA) which is the metabolite of vitamin A is essential for the migration of cells towards gut. After ingestion, retinal esters of vitamin A enter the liver and after being hydrolyzed into retinol subsequently bile gets enriched in retinol (Harrison and Hussain 2001, Wolf 2007). Inside the cell, alcohol dehydrogenase enzyme oxidizes retinol to retinal which is further oxidized by retinal dehydrogenase (RALDH) to RA (Mic, Molotkov et al. 2003). RA can bind to nuclear receptors including retinoic acid receptors and retinoid X receptors (Schug, Berry et al. 2007). Unlike DCs from other tissues, DCs in the GALT, express RALDH and thus induces the expression of gut homing receptor α4β7 and chemokine receptor CCR9 (Iwata, Hirakiyama et al. 2004, Hall, Grainger et al. 2011).

Thus, IgA is essential to recognize a wide variety of antigens at mucosal sites via diversified BCRs generated by CSR, SHM and receptor revision and subsequent clonal expansion (Spencer, Barone et al. 2009).

1.9 Aims

The specific aims of each chapter have been mentioned in the beginning of each section but succinctly can be stated as follows:

1- To determine how various biases could be observed in the ratios of in-frame to out-of-frame gene rearrangements at the *IGL* locus and their significance in studying biases at various stages of B cell development.

2- To determine how the repertoire of naïve B cells is forged at the *IGK* locus with an emphasis on various gene rearrangements that are selected for or against and consequences of retention of multiple non-functional gene rearrangements in the DNA.

3- To determine if there is any preference for light chains to select heavy chain variable genes during establishment of the repertoire of naïve B cells?

Chapter 2

Materials and methods

2.1 Collection of blood samples and PBMC isolation

The healthy buffy coats were purchased from the Blood Transfusion Service, Tooting, UK. Equal volumes of blood samples were mixed with RPMI medium (1:1). Fifteen ml of Ficoll-Hypaque was pipetted into 50ml falcon tubes. Approximately 35ml of diluted blood was layered over Ficoll-Hypaque solution. Tubes were centrifuged for 25 minutes at 400G with slow acceleration and without brakes. Peripheral blood mononuclear cells (PBMCs) were removed from the interface of Ficoll and plasma and washed in RPMI medium supplemented with 10% FCS at 4ºC for 5 minutes at 1200rpm. The supernatant was removed and cell yield was determined using haemocytometer. Cells not to be used immediately were resupended in FCS supplemented with 10% DMSO (freezing medium) at an approximate volume for storage in liquid nitrogen in 2ml cryovials at an approximate maximum cell density of 1 $X10^{10}$ cells/ml.

2.2 Working Solutions

2.2.1 Solutions for high throughput sequencing

2.2.1.1 Sort lysis reverse transcriptase (SLyRT) buffer

SLyRT buffer was used to synthesize cDNA after sorting of B cells. The recipe and concentrations of various ingredients used for making SLyRT buffer are given in the Table 2-1. The buffer was stored at -20ºC.

Table 2-1: Concentration and volume of various reagents used for making SLyRT buffer

Reagents	Initial concentration	For 1 reaction volume (µl)	For 50 reaction volume (µl)	Final volume concentration
First strand RT buffer (Invitrogen)	5X	8	400	1X
Random hexamers (QIAGEN)	50 ng/µl	12	600	15 ng/µl
Triton X-100 (Sigma)	5%	1	50	0.13% (v/v)
RiboSafe RNase inhibitor (Bioline)	40 U/µl	2.5	125	2.5 U/µl
DTT (Invitrogen)	0.1 M	4.5	225	11.25 mM
dNTP mix (Promega)	10 mM each	2	100	500 µM
PCR graded water	NA	6	300	NA

2.2.2 Solutions for agarose gel electrophoresis

2.2.2.1 50X Tris acetate EDTA (TAE) buffer

In order to prepare stock solution of 50X TAE buffer, 242g TRIS base (Sigma), 57.1ml glacial acetic acid (Sigma) and 100ml of 0.5M EDTA (Sigma) were dissolved in 800ml of distilled water and the volume was adjusted to IL following complete dissolution of chemicals. For experiments 1X buffer was prepared from this stock solution.

2.2.2.2 10X Tris borate EDTA (TBE) buffer

10X TBE buffer was prepared by mixing 108g Trizma base (Sigma), 55g boric acid (Sigma) and 8.3g EDTA (Sigma) were dissolved in 800ml of distilled water and the final volume was adjusted to IL.

2.2.2.3 Ethidium bromide stock (0.5 µl/ml)

10mg ethidium bromide was dissolved in 1ml distilled water and was diluted 1:20 with 1X TBE to make up a working solution of 50ng/100ml.

2.2.2.4 Gel loading buffer

Gel loading buffer was prepared by dissolving 40g sucrose, 0.25g bromophemol blue and 0.25g xylene cyanol in 100ml double distilled water.

2.2.3 Solutions for flow cytometry

2.2.3.1 10X Phosphate buffered saline

Stock solution of 10X phosphate buffered saline (PBS) was prepared by mixing 80g of NaCl, 2g of KCl, 14.4g of Na_2HPO_4 and 2.4g of KH_2PO_4 into 800ml of distilled water. After adjusting pH to 7.4 the final volume was made to IL with distilled water. The final solution was sterilized by autoclaving for 20 minutes at 120ºC. For all experiments 1X PBS was prepared by diluting 10X PBS.

2.2.3.2 RPMI-1640 medium

RPMI-1640 medium containing 1% L-glutamine (Gibco) was supplemented with 1% PenStrep (Sigma), 1% amphotericin B and 10% heat inactivated (56ºC for 30 minutes) FCS.

2.2.3.3 FACS buffer

2% fetal calf serum (seraLab) was diluted in 1X PBS.

2.2.4 Solutions for cloning

2.2.4.1 SOC medium

SOC medium was prepared by adding 2.0g tryptone (DIFCO), 0.5g yeast extract (DIFCO), 1ml NaCl (58.44g/L) and 0.25ml 1M KCl (74.55g/L) in 97ml water. The solution was autoclaved at 120ºC for 20 minutes. After sterilization the solution was allowed to cool and 1ml 2M Mg^{++} stock (filter sterilized 1M $MgCl_2 6H_20$/1M $MgSO_4 7H_20$ and 1ml 2M sucrose) was added and volume was adjusted to 100ml with sterile water.

2.2.4.2 IPTG stock solution (0.1M)

Solution of IPTG (0.1 M) was prepared by adding 0.6g IPTG (Promega, UK) in 25ml distilled water. The solution was sterilized by passing through a 0.22µm filter using a syringe. The filtrate was stored in 1.5ml aliquots at -20ºC.

2.2.4.3 Ampicillin stock solution (50mg/ml)

The stock solution of ampicillin was prepared by adding 1g ampicillin (Sigma, UK) in 20ml distilled water. The solution was sterilized by passing through 0.22µm filter using a syringe. The filtrate was stored in 1.5ml aliquots at -20ºC.

2.2.4.4 LB agar

In order to prepare LB agar, 1g tryptone (DIFCO, UK), 0.5g yeast extract (DIFCO, UK) and 1.5g Bacto agar (BD, UK) were dissolved in 100ml distilled water. The solution was sterilized by autoclaving for 20 minutes at 120ºC. LB agar was allowed to cool for 15 minutes after autoclaving. Then 200µl 0.5 µl/ml ampicillin (Sigma), 200µl 0.5 µl/ml X-Gal (Promega) and 500µl IPTG were added. After mixing well, 20ml agar was pipetted in petri dishes.

2.3 Protocols

2.3.1 Amplification of *IGL* gene rearrangements

2.3.1.1 Family specific nested PCR

Family specific PCR was performed using two rounds of nested PCR. *IGL* gene rearrangements were amplified using family specific IGLV1 and IGLV2 primers and IGLJ primers (Table 2-2). The PCR was conducted in 50µl volume containing 100 ng of 5' family specific IGLV1 and IGLV2 primers and 100ng IGLJ primers. Ten microliters of the DNA sample was used for the first round and 2 µl of first-round PCR products for the second round. All reactions contained 200 µM of each dNTP and 2.5 mM $MgCl_2$ in TaqDNA polymerase 1X reaction buffer. After hot start at 95°C for 10 min, 1 U TaqDNA polymerase (Promega) was added to each reaction at 60°C. The PCR programme for different *IGLV* families has been shown in Tables 2-3 and 2-4. PCR products were separated in 3.5% agarose gel. The product bands at 350 bp were selected for cloning and sequencing.

Table 2-2: Sequences of the primers used for family specific amplification of *IGLV*1 and *IGLV*2 families

Primer	**Sequence** (5' to 3')
Primers used in first round	
LJE	ACCMAGSTSACCGTCCT
VL1E	CCTGGGCCCAGTCTGTG
VL2E	CTCCTCA(GC)(TC)CTCCTCACT
Primers used in second round	
JL1N	GGMASTGGSACCAAGGTSACC
JL23N	GGMGGAGGSACCMAGCTGACC
VL1N	CCAGTCTGTG(TC)TGAC(TG)CAGCC
VL2N	CAGTCTGCCCTGACTCAGCC

Table 2-3: The PCR programme used for the amplification of the gene rearrangements involving *IGLV1* gene family

Step	Temperature (ºC)	Time	Cycles
Denaturation	95	30 s	35
Annealing	52.5 (first round) 50.5 (second round)	30 s	
Elongation	72	1.5 min	
Final extension	72	5 min	

Table 2-4: The PCR programme used for the amplification of the gene rearrangements involving *IGLV2* gene family

Step	Temperature (ºC)	Time	Cycles
Denaturation	95	30 s	35
Annealing	51 (first round) 55 (second round)	30 s	
Elongation	72	1.5 min	
Final extension	72	5 min	

2.3.1.2 Gel electrophoresis of amplified products

PCR products were separated in 3.5% agarose gel using 1X TBE buffer. The band of product size approximately 350bp was selected from each sample for further gene cloning and sequencing.

2.3.1.3 Purification of PCR products

For the purification of the amplified products QIAquick PCR purification kit was used. According to the manufacturer's protocol, 5 volumes of buffer (PB) was added to one volume of the PCR reaction and mixed. To bind DNA, the column (QIAquick) was placed in a 2ml collection tube and sample was passed by centrifuging at 13,000rpm for 1 minute. Flow through was discarded and the column was placed in the same collection tube. To wash, 750µl buffer (PE) was added in the column and passed by centrifuging at 13,000rpm for 1 minute. Flow through was discarded and the column was placed in the same collection tube.

The column was centrifuged once more for 1 minute to remove any residual wash buffer. In the next step, the column was placed in a clean eppendorf. To elute DNA, 30µl buffer (EB) was added in the centre of the column and centrifuged after 1 minute at 13,000rp for 1 minute.

2.3.2 Molecular cloning of PCR products

2.3.2.1 Ligation

The pGEM-T vector kit (Promega) was used to ligate PCR products. The final ligation reaction volume was adjusted to 10µl: 5µl ligation buffer (10X), 1µl of T4 DNA ligase 1µl vector and 3µl of purified PCR product. All chemicals were kept on ice during reaction. The final ligation mixture was kept at room temperature for 2 hours or 4ºC for overnight incubation.

2.3.2.2 Transformation

For transformation, 2.5µl of each ligation reaction mixture was mixed 25µl of competent cells of bacteria (JM109, Promega) in 1.5ml sterilized eppendorfs and incubated for 20 minutes on ice. The mixture was given a heat shock in a water bath at 42ºC for 45 seconds and subsequently incubated on ice for 2 minutes. Nine hundred and fifty microliters of SOC medium was added to each tube followed by shaking incubation at 37ºC for 2 hours at 150rpm. Bacterial cells were collected by centrifugation at 200rpm at the room temperature for 10 minutes and supernatant was discarded. The bacterial pellet was re-suspended in 100µl of ice cold SOC medium and spread on LB agar plates supplemented with antibiotics, IPTG and X-Gal. Plates were inverted and placed in an incubator at 37ºC for overnight.

2.3.2.3 Blue/white clone screening

The bacteria successfully transformed with pGEM-T vector containing insert of *IGLV-IGLJ* gene rearrangement grew as white colonies. Bacteria that failed to be transformed grew as blue colonies. White colonies were picked with sterile pipette tips and transferred to gridded plate and allowed to grow overnight at 37ºC. Pipette tips used to select colonies were rinsed in 15µl water in individual wells of a 96 well plate. The plates were covered with sticking seal and placed in a PCR machine at 95ºC for 10 minutes. The plate was centrifuged for 10

minutes at 400g for 10 minutes and stored at -20ºC. The cell debris settled at the bottom of the plate and the supernatant water was used as a source of cloned DNA fragment, to be amplified in the next step of clone screening as mentioned below.

2.3.2.4 Clone screening by PCR

For screening of clones 5µl of the supernatant was used as the template for a PCR reaction. The PCR reaction mixture included 1.5mM $MgCl_2$, 0.2M dNTPs, 1X Taq DNA polymerase reaction buffer, 50ng of each M13 forward and M13 reverse primers (Table 2.4) and 0.25U of Taq polymerase in a total volume of 10µl. The sequences of primers and conditions for amplification are given in the Tables 2-5 and 2-6.

Table 2-5: Sequences of M13 primers used for screening of transformed clones

Primer	Sequence 5' to 3'
M13 forward	GTAAAACGACGGCCAGT
M13 reverse	GGAAACAGCTATGACCATG

Table 2-6: PCR conditions used for screening of clones to detect desired insert of rearranged immunoglobulin lambda light chain genes

Stage	Temperature (ºC)	Time	Cycles
Hot start	95	5 minutes	
Denaturation	95	30 seconds	30
Annealing	45	30 seconds	
Elongation	72	5 minutes	

2.3.2.5 DNA sequencing

Clones were selected from gridded plate according to the size of the cloned PCR products and transferred to individual wells of a 96 well plate containing LB agar supplemented with 200µl ampicillin (0.5µl/ml). Plates were incubated at 37ºC overnight before sending to Beckman and Coulter for sequencing.

2.3.3 Flow cytometry

PBMCs were thawed following liquid nitrogen storage. Cells were then washed in 15ml RPMI 1640 medium and centrifuged at 1200rpm for 5 minutes. Cells were counted and immunostained in 100µl 2% FACS buffer with the volumes of antibodies as listed in the Tables 2-7 and 2-8. The optimal volume of antibodies for usage had previously been determined by titration. For staining cells were incubated for 20 minutes on ice in the dark. After staining 2ml 2% FACS buffer was added and centrifuged at 1200rpm for 5 minutes at 4ºC. Cells were resuspended in 100µl 1XPBS and stained with 1µl of viability stain for 30 minutes. After staining cells were washed with 2ml of 2% FACS buffer. Cells were resuspended in 1500µl of 2% FACS buffer.

For each experiment following controls were also prepared:

1. **Single stain control:** Negative and positive beads (1:1) were stained for single antibody in 2% FACS buffer for 20 minutes on ice.
2. **FMO control:** In order to determine the background stain noise, FMO (fluorescence minus one) controls were prepared for each flurochrome used in a particular experiment. The FMO control was prepared by staining a few cells in 2% FACS buffer on ice for 20 minutes with all fluorescence tagged monoclonal antibodies excluding one in the respective tube.
3. **Isotype control:** The isotype control was prepared to determine the level of non specific binding. For this purpose, a few cells were stained in 2% FACS buffer for 20 minutes on ice with control isotype tagged with the flurochrome used in a particular experiment.

Cells for sorting were sorted into polypropylene FACS tubes containing 1.5ml of 2% FACS buffer. Upon completion of cell sorting, cells were centrifuged for 5 minutes at 300g and supernatant was removed. For flow cytometric analysis without sorting, cells were acquired and recorded using the FACSCanto (BD). Cell sorting was performed using the FACSAria-II (BD) with sort purities >90%. All flow cytometric experiments could be subsequently analyzed on FlowJo (V7.5).

Table 2-7: Surface markers considered for the characterization of various B cell subsets

B cell subset	Phenotypic definition
Transitional B cells	$CD19^{+}IgD^{+}CD27^{-}CD38^{++}$ or $CD19^{+}IgD^{+}CD27^{-}CD10^{++}$
Mature naïve B cells	$CD19^{+}IgD^{+}CD27^{-}CD38^{+/-}$ or $CD19^{+}IgD^{+}CD27^{-}CD10^{-}$
IgA memory	$CD19^{+}CD27^{+}CD20^{+}IgA^{+}$
IgA plasmablast	$CD19^{+}CD27^{+}CD20^{-}IgA^{+}$ or $CD19^{+}CD27^{+}CD38^{++}IgA^{+}$
$CD27^{+}$ IgA	$CD19^{+}CD27^{+}CD38^{+/-}IgA^{+}$
$CD27^{-}$ IgA	$CD19^{+}CD27^{-}CD38^{+/-}IgA^{+}$
Mature naïve expressing kappa light chain	$CD19^{+}IgD^{+}CD27^{-}CD38^{+/-}IGK^{+}$
Mature naïve expressing lambda light chain	$CD19^{+}IgD^{+}CD27^{-}CD38^{+/-}IGL^{+}$

Table 2-8: Mouse anti human antibodies used for flow cytometric analysis and sorting

Mouse anti human antibodies	Fluorochrome	Volume in 100 µl of 2% FACS buffer	Maximum number of cells/stain volume	Isotypes	Manufacturer
CD19	PerCp-Cy5.5	5µl/million cells	<10 million	Mouse IgG1 k PerCP-cy5.5	BD Pharm
IgD	PE	5µl/million cells		Mouse IgG1 k PE	Biolegend
CD27	APC	5µl/million cells		Mouse IgG1 k APC	BD Pharmingen
CD10	APC	5µl/million cells		Mouse IgG1 k APC	Biolegend
CD38	PE	5µl/million cells		Mouse IgG1 k PE	eBioscience
IgA	FITC	10µl/million cells		Mouse IgG1 k FITC	Miltenyi Biotech
IgK	APC-H7	5µl/million cells		Mouse IgG1 k APC-H7	BD Pharm
IgL	Pacific blue	3µl/million cells		Mouse IgG1 k pacific blue	BioLegend
Viability stain	Live/Dead Fixable aqua	1µl/million cells	N.A.		Invitrogen

2.3.4 High throughput sequencing

2.3.4.1 cDNA synthesis

The overall process of HTS is summarized in Figure 2-1. To synthesize cDNA from sorted populations, cells were centrifuged at 1200rpm for 5 minutes and the supernatant was removed. Cells were resuspended in 180µl sort lysis reverse transcriptase (SLyRT) buffer. Tubes were inverted several times to ensure immersion of cells in the buffer and then centrifuged at 3000rpm for 2 minutes to collect the contents. Twenty microliters of 500U (1:8) of SuperScript III reverse transcriptase (RT; Invitrogen) was added. The process of reverse transcription was carried out in a thermal cycler according to conditions given in Table 2-9.

Table 2-9: Thermal conditions for reverse transcription

Steps	Temperature (ºC)	Time (minutes)
Denaturation	42	10
Annealing	25	10
RT (extension)	55	60
RT (termination)	72	15

PCR tubes were centrifuged at 13,000rpm for 5 minutes after thermal cycling. The top fraction of supernatant was transferred to clean eppendorfs immediately without disturbing the debris. Samples were stored at -20ºC.

2.3.4.2 Genomic DNA isolation

The genomic DNA was isolated from FACS sorted populations of B cells using DNeasy blood and tissue kit (Qiagen). According to the manufacturer's protocol after FACS sorting cells were centrifuged for 5 minutes at 300g and resuspended in 200µl 1X PBS. After adding 20µl of proteinase K and 200µl of lysis Buffer (AL) cells were vortexed and incubated at 56ºC for 20 minutes. Two hundred microliters of ethanol (96% to 100%) was added and cells were mixed thoroughly by vortexing. This mixture was passed through spin columns fitted in collection tubes by centrifugation at 8,000rpm for 1 minute. The collection tubes filled with flow through were discarded. Spin columns were placed in new collection tubes and 500µl washing buffer (AW1) was added and centrifuged at 8,000rpm for 1 minute. Collection tubes

containing flow through were discarded. Spin columns were placed in new collection tubes and 500µl washing buffer (AW2) was added followed by centrifugation at 14,000rpm for 3 minutes. Collection tubes containing flow through were discarded and spin columns were placed in 1.5ml clean eppendorfs. Two hundred microliters of elution buffer AE was added followed by centrifugation at 8,000rpm for 1 minute. Samples of eluted DNA were stored at -20ºC.

2.3.4.3 PCR amplification for high throughput sequencing

High throughput sequencing PCR amplification of genomic and cDNA was performed in two rounds, where the same primers were used in both reactions, though the second PCR primers were MID (multiplex identifier) tagged. Genomic (*IGKV*-I*GKJ*) and cDNA (*IGKV*-*IGKC*) gene rearrangements were amplified using semi nested PCR. In PCR1, 25µl reaction mixture contained 6.25µl of cDNA, 0.625U Phusion DNA polymerase, 200µM of each dNTPs, 41.75nM of each upstream *IGKV* primers, and 250nM downstream IGKC primer in 1 X reaction buffer. After a hot start at 98ºC (for 30 seconds, hold at 50ºC) Phusion DNA polymerase was added, followed by 15 cycles of 98ºC (10 seconds), 58ºC (15 seconds) and 72ºC (30 seconds) and 1 cycle of 72ºC (5 minutes).

In order to produce sufficient sample for high throughput sequencing, amplified product from the first round of PCR (PCR1) was re-amplified (PCR2). For the second round of PCR, the same primers were used as in the first round but these were modified by adding a bar code consisting of 10 nucleotides called multiplex identifier (MID) at the end of the primer sequence. The MID nucleotide sequences were designed based on the criterion that these should not have binding affinity anywhere in the genome. The MIDs only acted as a barcode and were specific for each B cell subset and individual studied. MIDs were added to make experiments cost effective as this approach enabled pooling of multiple samples into 1 sequencing sample and segregation of sequences after sequencing.

Twenty microliters of PCR2 reaction mixture contained 2µl of PCR1 product, 0.5U Phusion DNA polymerase, 200µM each dNTPs, 41.75nM each upstream MID: *IGKV*1-6 primers, 250nM *IGKC* downstream primer in 1X reaction buffer. PCR2 was performed with conditions: 98ºC (30 seconds), 20 cycles of 98ºC (10 seconds), 58ºC (15 seconds), 72ºC

(30 seconds) and 1 cycle of 72ºC (5 minutes). Recipes and summaries for each PCR reaction are given in Table 2-10 and 2-11.

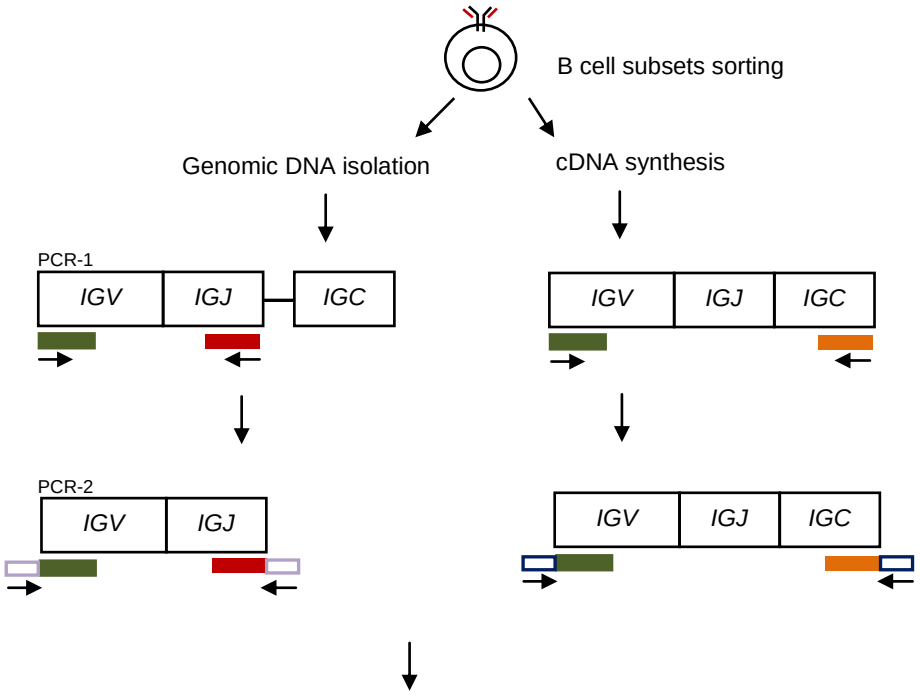


Figure 2-1: Schematic representation of the steps involved in the experiment of high throughput sequencing. B cell subsets were isolated by flow cytometry using specific monoclonal antibodies against cell surface markers. Gene rearrangements were amplified using genomic DNA (*IGV-IGJ*) or cDNA (*IGV-IGJ)* by PCR. Amplification involved two rounds (PCR1 and PCR2) to generate abundant sample for HTS sequencing. Because in the final step, amplified products from all individuals were pooled in one sample, MID tagged primers (purple and blue open rectangles) were used in the second round of PCR to distinguish various products post sequencing. MID barcodes were specific for each individual and B cell subset. After PCR-amplification, desired product bands were cut from the gel, purified, concentrated and sent for sequencing and analyzed subsequently.

Table 2-10: Recipes and concentrations for HTS first round PCR reactions (primers without MID tags)

Reagents	Stock solution	µl/reaction	Final concentration
GC buffer	5X	5	1X
1:40 Phusion DNA polymerase	0.125U/µl	5	0.625 U
dNTPs	20mM each	0.25	200µM each
*IGKV*1-6 (5') primer mix *IGKC* (3') primer mix	835nM 5µM	1.25	41.75nM each + 250nM
PCR water	-	7.25	-
cDNA	-	6.25	~200ng
Volume	-	25	

Table 2-11: Recipes and concentrations for HTS second round PCR reactions (primers with MID tags)

Reagents	Stock solution	µl/reaction	Final concentration
GC buffer	5X	4	1X
Phusion DNA polymerase 1:20	0.125U/ µl	2	0.5 U
dNTPs	20mM each	0.2	200µM each
MID: *IGKV*1-6 (5') primer mix MID: *IGKC* (3') primer mix	835nM 5µM	1	41.75nM each + 250nM
PCR water	-	10.8	-
cDNA	-	2	
Volume	-	20	

In order to obtain sufficient quantities of PCR product for sequencing first round of PCR reactions were performed eight times for each DNA sample from which second round of PCR reactions were carried out in duplicate (final volume per DNA sample was 20µl X 16 = 320µl). Total PCR products from 16 second round PCRs of the same MID tag were pooled and heated to 95ºC for 5 minutes. Sequences of various PCR primers used for the amplification of rearranged *IGH*, *IGK* and *IGL* loci are given in Tables (2-12 to 2-17).

2.3.4.4 Gel electrophoresis of MID tagged PCR products, purification, pooling and enrichment

For running PCR products of high throughput sequencing PCR, 1X TAE buffer was used to prepare 3.5% agarose gel using 1X TAE buffer. PCR products were run for 40 minutes at 120V and bands of target were size extracted and purified using QIAquick Gel Purification Kit (Qiagen). According to manufacturer's protocol gel slices were incubated with 5 volumes of Buffer (QG) per amplicons type in 15ml falcon tubes at 55ºC and vortexed frequently until gel slices were melted. One QIAquick spin column was used for each amplicons and 750µl sample solution was passed through column by centrifugation at 13,000 rpm for 30 seconds. Flow through was collected in another falcon tube. This process was repeated until whole solution was passed through the column. The flow through that was collected in another tube was passed through the column again. After passing the whole flow through, the spin column was washed with 750µl buffer (PE) by centrifugation at 13,000 rpm for 1 minute. For eluting DNA from the column 55µl PCR grade water was added on top of the column and allowed to stand for 3 minutes at room temperature. The column was centrifuged at 13,000 rpm for 1 minute and eluted solution was collected in 1.5ml microcentrifuge tube. Additional 55µl of PCR grade water was added and again allowed to stand for 3 minutes at room temperature. The column was spun at 13,000rpm for 1 minute and eluted solution was collected in same 1.5ml microcentrifuge tube. The concentration of eluted DNA was estimated using the Qubit dsDNA HS Assay kit on the Qubit 2.0 Fluorometer. After determining the concentration of each sample, equal quantities of DNA samples were pooled. The pooled products were enriched using QIAquick PCR purification kit. The methodology was same as used for purification of DNA products excised from agarose gel (described above).

2.3.4.5 Sequence analysis

The sequencing data was returned in .Fna format files which were then put through an automated analysis pipeline by Dr. David Kipling (Cardiff University) and Dr. Deborah Dunn-Walters (King's College London). The data was subjected to a stringent set of rules designed to quality control the data.

Sequences were assigned to the corresponding samples based on the terminal MID sequence. Sequences that contained a second MID sequence that was either different or located internally were excluded. A series of stringent quality control criteria were applied to exclude biologically implausible sequences. *IGKV-IGKC* rearrangements amplified from cDNA were only accepted as biologically plausible if they were over 415 nucleotides in length and *IGKV-IGKJ* rearrangements if they were over 290 nucleotides in length. Sequences that failed to start with a MID tag were rejected. After the initial quality control steps, sequences underwent immunoglobulin genotyping using the IMGT (Immunoglobulin immunogenetics) V-Quest (http://www.imgt.org/IMGT_vquest/share/textes/) programme. The IMGT directory was developed in 1989 by Marie-Paule Lefranc. It provides various databases and tools to analyze the BCR and TCR repertoires (Lefranc, Giudicelli et al. 2015). In the next step, clonotype clustering was done using a clustering matrix and one modal sequence was selected as a unique sequence to address the questions being asked in this thesis. The Protpram program was used to calculate various physicochemical properties of CDR-3 amino acid sequences (http://web.expasy.org/protparam/) which includes molecular weight, theoretical PI, aliphatic index, GRAVY index and amino acid composition.

The Excel spread sheets were retrieved containing information of immunoglobulin genes, clones and physical characteristics. Data was analyzed using pivot tables. Gene frequencies were compared using X^2 tests with Bonferroni correction for multiple comparisons and the significance of the differences in the physical properties was determined by employing Student's t-test.

Table 2-12: Sequences of primers used for the amplification of rearranged lambda light chain genes during first round of high throughput sequencing PCR

Primers	**Sequence** (5' to 3')
IGLV1	CAGTCTGTGCTGACKCAGCC
IGLV2	CAGTCTGCCCTGACTCAGCC
IGLV3	CCTATGAGCTGACWCAGCCAC
IGLV4/5	CAGCCTGTGCTGACTCARYC
IGLV6	CCAGNCTGTGSTGACTCAG
IGLJE	ACCMAGSTSACCGTCCT

Table 2-13: Sequences of MID-tagged primers used for the amplification of rearranged lambda light chain genes during second round of high throughput sequencing PCR

Primer	**Sequence** (5' to 3')	**Primer**	**Sequence** (5' to 3')	**Primer**	**Sequence** (5' to 3')
MID1 IGLV1	acgagtgcgtCAGTCTGTGCTGACKCAGCC	MID5 IGLV1	atcagacacgCAGTCTGTGCTGACKCAGCC	MID9 IGLV1	tagtatcagcCAGTCTGTGCTGACKCAGCC
MID1 IGLV2	acgagtgcgtCAGTCTGCCCTGACTCAGCC	MID5 IGLV2	atcagacacgCAGTCTGCCCTGACTCAGCC	MID9 IGLV2	tagtatcagcCAGTCTGCCCTGACTCAGCC
MID1 IGLV3	acgagtgcgtCCTATGAGCTGACWCAGCCAC	MID5 IGLV3	atcagacacgCCTATGAGCTGACWCAGCCAC	MID9 IGLV3	tagtatcagcCCTATGAGCTGACWCAGCCAC
MID1 IGLV4/5	acgagtgcgtCAGCCTGTGCTGACTCARYC	MID5 IGLV4/5	atcagacacgCAGCCTGTGCTGACTCARYC	MID9 IGLV4/5	tagtatcagcCAGCCTGTGCTGACTCARYC
MID1 IGLV6	acgagtgcgtCCAGNCTGTGSTGACTCAG	MID5 IGLV6	atcagacacgCCAGNCTGTGSTGACTCAG	MID9 IGLV6	tagtatcagcCCAGNCTGTGSTGACTCAG
MID1 IGLJ 1N	acgagtgcgt GGMASTGGSACCAAGGTSACC	MID5 IGLJ 1N	atcagacacg GGMASTGGSACCAAGGTSACC	MID9 IGLJ 1N	tagtatcagc GGMASTGGSACCAAGGTSACC
MID1 IGLJ 23N	acgagtgcgt GGMGGAGGSACCMAGCTGACC	MID5 IGLJ 23N	atcagacacg GGMGGAGGSACCMAGCTGACC	MID9 IGLJ 23N	tagtatcagc GGMGGAGGSACCMAGCTGACC
MID2 IGLV1	acgctcgacaCAGTCTGTGCTGACKCAGCC	MID6 IGLV1	atatcgcgagCAGTCTGTGCTGACKCAGCC	MID10 IGLV1	tctctatgcgCAGTCTGTGCTGACKCAGCC
MID2 IGLV2	acgctcgacaCAGTCTGCCCTGACTCAGCC	MID6 IGLV2	atatcgcgagCAGTCTGCCCTGACTCAGCC	MID10 IGLV2	tctctatgcgCAGTCTGCCCTGACTCAGCC
MID2 IGLV3	acgctcgacaCCTATGAGCTGACWCAGCCAC	MID6 IGLV3	atatcgcgagCCTATGAGCTGACWCAGCCAC	MID10 IGLV3	tctctatgcgCCTATGAGCTGACWCAGCCAC
MID2 IGLV4/5	acgctcgacaCAGCCTGTGCTGACTCARYC	MID6 IGLV4/5	atatcgcgagCAGCCTGTGCTGACTCARYC	MID10 IGLV4/5	tctctatgcgCAGCCTGTGCTGACTCARYC
MID2 IGLV6	acgctcgacaCCAGNCTGTGSTGACTCAG	MID6 IGLV6	atatcgcgagCCAGNCTGTGSTGACTCAG	MID10 IGLV6	tctctatgcgCCAGNCTGTGSTGACTCAG
MID2 IGLJ 1N	acgctcgaca GGMASTGGSACCAAGGTSACC	MID6 IGLJ 1N	atatcgcgag GGMASTGGSACCAAGGTSACC	MID10 IGLJ 1N	tctctatgcg GGMASTGGSACCAAGGTSACC
MID2 IGLJ 23N	acgctcgaca GGMGGAGGSACCMAGCTGACC	MID6 IGLJ 23N	atatcgcgag GGMGGAGGSACCMAGCTGACC	MID10 IGLJ 23N	tctctatgcg GGMGGAGGSACCMAGCTGACC
MID3 IGLV1	agacgcactcCAGTCTGTGCTGACKCAGCC	MID7 IGLV1	cgtgtctctaCAGTCTGTGCTGACKCAGCC	MID11 IGLV1	tgatacgtctCAGTCTGTGCTGACKCAGCC
MID3 IGLV2	agacgcactcCAGTCTGCCCTGACTCAGCC	MID7 IGLV2	cgtgtctctaCAGTCTGCCCTGACTCAGCC	MID11 IGLV2	tgatacgtctCAGTCTGCCCTGACTCAGCC
MID3 IGLV3	agacgcactcCCTATGAGCTGACWCAGCCAC	MID7 IGLV3	cgtgtctctaCCTATGAGCTGACWCAGCCAC	MID11 IGLV3	tgatacgtctCCTATGAGCTGACWCAGCCAC
MID3 IGLV4/5	agacgcactcCAGCCTGTGCTGACTCARYC	MID7 IGLV4/5	cgtgtctctaCAGCCTGTGCTGACTCARYC	MID11 IGLV4/5	tgatacgtctCAGCCTGTGCTGACTCARYC
MID3 IGLV6	agacgcactcCCAGNCTGTGSTGACTCAG	MID7 IGLV6	cgtgtctctaCCAGNCTGTGSTGACTCAG	MID11 IGLV6	tgatacgtctCCAGNCTGTGSTGACTCAG
MID3 IGLJ 1N	agacgcactc GGMASTGGSACCAAGGTSACC	MID7 IGLJ 1N	cgtgtctcta GGMASTGGSACCAAGGTSACC	MID11 IGLJ 1N	tgatacgtct GGMASTGGSACCAAGGTSACC
MID3 IGLJ 23N	agacgcactc GGMGGAGGSACCMAGCTGACC	MID7 IGLJ 23N	cgtgtctcta GGMGGAGGSACCMAGCTGACC	MID11 IGLJ 23N	tgatacgtct GGMGGAGGSACCMAGCTGACC
MID4 IGLV1	agcactgtagCAGTCTGTGCTGACKCAGCC	MID8 IGLV1	ctcgcgtgtcCAGTCTGTGCTGACKCAGCC	MID12 IGLV1	tactgagctaCAGTCTGTGCTGACKCAGCC
MID4 IGLV2	agcactgtagCAGTCTGCCCTGACTCAGCC	MID8 IGLV2	ctcgcgtgtcCAGTCTGCCCTGACTCAGCC	MID12 IGLV2	tactgagctaCAGTCTGCCCTGACTCAGCC
MID4 IGLV3	agcactgtagCCTATGAGCTGACWCAGCCAC	MID8 IGLV3	ctcgcgtgtcCCTATGAGCTGACWCAGCCAC	MID12 IGLV3	tactgagctaCCTATGAGCTGACWCAGCCAC
MID4 IGLV4/5	agcactgtagCAGCCTGTGCTGACTCARYC	MID8 IGLV4/5	ctcgcgtgtcCAGCCTGTGCTGACTCARYC	MID12 IGLV4/5	tactgagctaCAGCCTGTGCTGACTCARYC
MID4 IGLV6	agcactgtagCCAGNCTGTGSTGACTCAG	MID8 IGLV6	ctcgcgtgtcCCAGNCTGTGSTGACTCAG	MID12 IGLV6	tactgagctaCCAGNCTGTGSTGACTCAG
MID4 IGLJ 1N	agcactgtag GGMASTGGSACCAAGGTSACC	MID8 IGLJ 1N	ctcgcgtgtc GGMASTGGSACCAAGGTSACC	MID12 IGLJ 1N	tactgagcta GGMASTGGSACCAAGGTSACC
MID4 IGLJ 23N	agcactgtag GGMGGAGGSACCMAGCTGACC	MID8 IGLJ 23N	ctcgcgtgtc GGMGGAGGSACCMAGCTGACC	MID12 IGLJ 23N	tactgagcta GGMGGAGGSACCMAGCTGACC

Table 2-14: Sequences of primers used for the amplification of rearranged heavy chain genes during first round of high throughput sequencing PCR

Primer	**Sequence** (5' to 3')
IGHV1	CCTCAGTGAAGGTCTCCTGCAAGG
IGHV2	TCCTGCGCTGGTGAAACCCACACA
IGHV3	GGTCCCTGAGACTCTCCTGTGCA
IGHV4	TCGGAGACCCTGTCCCTCACCTGC
IGHV5	CAGTCTGGAGCAGAGGTGAAA
IGHV6	CCTGTGCCATCTCCGGGGACAGTG
IGHM	GGGGAATTCTCACAGGAGAC

Table 2-15: Sequences of MID-tagged primers used for the amplification of rearranged heavy chain genes during second round of high throughput sequencing PCR

Primer	**Sequence** (5' to 3')	**Primer**	**Sequence** (5' to 3')
MID3 IGHV1	agacgcactcCCTCAGTGAAGGTCTCCTGCAAGG	MID8 IGHV1	ctcgcgtgtcCCTCAGTGAAGGTCTCCTGCAAGG
MID3 IGHV2	agacgcactcTCCTGCGCTGGTGAAACCCACACA	MID8 IGHV2	ctcgcgtgtcTCCTGCGCTGGTGAAACCCACACA
MID3 IGHV3	agacgcactcGGTCCCTGAGACTCTCCTGTGCA	MID8 IGHV3	ctcgcgtgtcGGTCCCTGAGACTCTCCTGTGCA
MID3 IGHV4	agacgcactcTCGGAGACCCTGTCCCTCACCTGC	MID8 IGHV4	ctcgcgtgtcTCGGAGACCCTGTCCCTCACCTGC
MID3 IGHV5	agacgcactcCAGTCTGGAGCAGAGGTGAAA	MID8 IGHV5	ctcgcgtgtcCAGTCTGGAGCAGAGGTGAAA
MID3 IGHV6	agacgcactcCCTGTGCCATCTCCGGGGACAGTG	MID8 IGHV6	ctcgcgtgtcCCTGTGCCATCTCCGGGGACAGTG
MID3 IGHM	agacgcactcGGGGAATTCTCACAGGAGAC	MID8 IGHM	ctcgcgtgtcGGGGAATTCTCACAGGAGAC
MID4 IGHV1	agcactgtagCTCAGTGAAGGTCTCCTGCAAGG	MID11 IGHV1	tgatacgtctCCTCAGTGAAGGTCTCCTGCAAGG
MID4 IGHV2	agcactgtagCCTGCGCTGGTGAAACCCACACA	MID11 IGHV2	tgatacgtctTCCTGCGCTGGTGAAACCCACACA
MID4 IGHV3	agcactgtagGGTCCCTGAGACTCTCCTGTGCA	MID11 IGHV3	tgatacgtctGGTCCCTGAGACTCTCCTGTGCA
MID4 IGHV4	agcactgtagTCGGAGACCCTGTCCCTCACCTGC	MID11 IGHV4	tgatacgtctTCGGAGACCCTGTCCCTCACCTGC
MID4 IGHV5	agcactgtagCAGTCTGGAGCAGAGGTGAAA	MID11 IGHV5	tgatacgtctCAGTCTGGAGCAGAGGTGAAA
MID4 IGHV6	agcactgtagCCTGTGCCATCTCCGGGGACAGTG	MID11 IGHV6	tgatacgtctCCTGTGCCATCTCCGGGGACAGTG
MID4 IGHM	agcactgtagGGGGAATTCTCACAGGAGAC	MID11 IGHM	tgatacgtctGGGGAATTCTCACAGGAGAC
MID7 IGHV1	cgtgtctctaCCTCAGTGAAGGTCTCCTGCAAGG	MID12 IGHV1	tactgagctaCCTCAGTGAAGGTCTCCTGCAAGG
MID7 IGHV2	cgtgtctctaTCCTGCGCTGGTGAAACCCACACA	MID12 IGHV2	tactgagctaTCCTGCGCTGGTGAAACCCACACA
MID7 IGHV3	cgtgtctctaGGTCCCTGAGACTCTCCTGTGCA	MID12 IGHV3	tactgagctaGGTCCCTGAGACTCTCCTGTGCA
MID7 IGHV4	cgtgtctctaTCGGAGACCCTGTCCCTCACCTGC	MID12 IGHV4	tactgagctaTCGGAGACCCTGTCCCTCACCTGC
MID7 IGHV5	cgtgtctctaCAGTCTGGAGCAGAGGTGAAA	MID12 IGHV5	tactgagctaCAGTCTGGAGCAGAGGTGAAA
MID7 IGHV6	cgtgtctctaCCTGTGCCATCTCCGGGGACAGTG	MID12 IGHV6	tactgagctaCCTGTGCCATCTCCGGGGACAGTG
MID7 IGHM	cgtgtctctaGGGGAATTCTCACAGGAGAC	MID12 IGHM	tactgagctaGGGGAATTCTCACAGGAGAC

Table 2-16: Sequences of primers used for the amplification of rearranged kappa light chain genes during first round of high throughput sequencing PCR

Primer	**Sequence** (5' to 3')
IGKV1	CATCCAGWTGACCCAGTCTCC
IGKV2	GATATTGTGATGACCCAGWCT
IGKV 3	GACRCAGTCTCCAGCCACCCTG
IGKV4	GACATCGTGATGACCCAGTCT
IGKV5	GAAACGACACTCACGCAGTCT
IGKV6	GAAATTGTGCTGACTCAGTCT
IGKJ1/4	TCCACCTTGGTCCCTYSGCCG
IGKJ2	CTCCAGCTTGGTCCCCTGGCCA
IGKJ3	ATCCACTTTGGTCCCAGGGCCG
IGKJ5	CTCCAGTCGTGTCCCTTGGCCG
IGKC	CCTTCCACTGTACTTTGGCCTC

Table 2-17: Sequences of MID-tagged primers used for the amplification of rearranged kappa light chain genes during second round of high throughput sequencing PCR

Primer	**Sequence** (5' to 3')	Primer	**Sequence** (5' to 3')	Primer	**Sequence** (5' to 3')
MID1 IGKV1	acgagtgcgtCATCCAGWTGACCCAGTCTCC	MID5 IGKV1	atcagacacgCATCCAGWTGACCCAGTCTCC	MID9 IGKV1	tagtatcagcCATCCAGWTGACCCAGTCTCC
MID1 IGKV2	acgagtgcgtGATATTGTGATGACCCAGWCT	MID5 IGKV2	atcagacacgGATATTGTGATGACCCAGWCT	MID9 IGKV2	tagtatcagcGATATTGTGATGACCCAGWCT
MID1 IGKV3	acgagtgcgtGACRCAGTCTCCAGCCACCCTG	MID5 IGKV3	atcagacacgGACRCAGTCTCCAGCCACCCTG	MID9 IGKV3	tagtatcagcGACRCAGTCTCCAGCCACCCTG
MID1 IGKV4	acgagtgcgtGACATCGTGATGACCCAGTCT	MID5 IGKV4	atcagacacgGACATCGTGATGACCCAGTCT	MID9 IGKV4	tagtatcagcGACATCGTGATGACCCAGTCT
MID1 IGKV5	acgagtgcgtGAAACGACACTCACGCAGTCT	MID5 IGKV5	atcagacacgGAAACGACACTCACGCAGTCT	MID9 IGKV5	tagtatcagcGAAACGACACTCACGCAGTCT
MID1 IGKV6	acgagtgcgtGAAATTGTGCTGACTCAGTCT	MID5 IGKV6	atcagacacgGAAATTGTGCTGACTCAGTCT	MID9 IGKV6	tagtatcagcGAAATTGTGCTGACTCAGTCT
MID1 IGKJ1/4	acgagtgcgtTCCACCTTGGTCCCTYSGCCG	MID5 IGKJ1/4	atcagacacgTCCACCTTGGTCCCTYSGCCG	MID9 IGKJ1/4	tagtatcagcTCCACCTTGGTCCCTYSGCCG
MID1 IGKJ2	acgagtgcgtCTCCAGCTTGGTCCCCTGGCCA	MID5 IGKJ2	atcagacacgCTCCAGCTTGGTCCCCTGGCCA	MID9 IGKJ3	tagtatcagcATCCACTTTGGTCCCAGGGCCG
MID1 IGKJ3	acgagtgcgtATCCACTTTGGTCCCAGGGCCG	MID5 IGKJ3	atcagacacgATCCACTTTGGTCCCAGGGCCG	MID9 IGKJ2	tagtatcagcCTCCAGCTTGGTCCCCTGGCCA
MID1 IGKJ5	acgagtgcgtCTCCAGTCGTGTCCCTTGGCCG	MID5 IGKJ5	atcagacacgCTCCAGTCGTGTCCCTTGGCCG	MID9 IGKJ5	tagtatcagcCTCCAGTCGTGTCCCTTGGCCG
MID1 IGKC	acgagtgcgtCCTTCCACTGTACTTTGGCCTC	MID5 IGKC	atcagacacgCCTTCCACTGTACTTTGGCCTC	MID9 IGKC	tagtatcagcCCTTCCACTGTACTTTGGCCTC
MID2 IGKV1	acgctcgacaCATCCAGWTGACCCAGTCTCC	MID6 IGKV1	atatcgcgagCATCCAGWTGACCCAGTCTCC	MID10 IGKV1	tgatacgtctCATCCAGWTGACCCAGTCTCC
MID2 IGKVK2	acgctcgacaGATATTGTGATGACCCAGWCT	MID6 IGKV2	atatcgcgagGATATTGTGATGACCCAGWCT	MID10 IGVK2	tgatacgtctGATATTGTGATGACCCAGWCT
MID2 IGKV3	acgctcgacaGACRCAGTCTCCAGCCACCCTG	MID6 IGKV3	atatcgcgagGACRCAGTCTCCAGCCACCCTG	MID10 IGVK3	tgatacgtctGACRCAGTCTCCAGCCACCCTG
MID2 IGKV4	acgctcgacaGACATCGTGATGACCCAGTCT	MID6 IGKV4	atatcgcgagGACATCGTGATGACCCAGTCT	MID10 IGVK4	tgatacgtctGACATCGTGATGACCCAGTCT
MID2 IGKV5	acgctcgacaGAAACGACACTCACGCAGTCT	MID6 IGKV5	atatcgcgagGAAACGACACTCACGCAGTCT	MID10 IGVK5	tgatacgtctGAAACGACACTCACGCAGTCT
MID2 IGKV6	acgctcgacaGAAATTGTGCTGACTCAGTCT	MID6 IGKV6	atatcgcgagGAAATTGTGCTGACTCAGTCT	MID10 IGVK6	tgatacgtctGAAATTGTGCTGACTCAGTCT
MID2 IGKJ1/4	acgctcgacaTCCACCTTGGTCCCTYSGCCG	MID6 IGKJ1/4	atatcgcgagTCCACCTTGGTCCCTYSGCCG	MID10 IGKJ1/4	tgatacgtctTCCACCTTGGTCCCTYSGCCG
MID2 IGKJ2	acgctcgacaCTCCAGCTTGGTCCCCTGGCCA	MID6 IGKJ2	atatcgcgagCTCCAGCTTGGTCCCCTGGCCA	MID10 IGKJ3	tgatacgtctATCCACTTTGGTCCCAGGGCCG
MID2 IGKJ3	acgctcgacaATCCACTTTGGTCCCAGGGCCG	MID6 IGKJ3	atatcgcgagATCCACTTTGGTCCCAGGGCCG	MID10 IGKJ2	tgatacgtctCTCCAGCTTGGTCCCCTGGCCA
MID2 IGKJ5	acgctcgacaCTCCAGTCGTGTCCCTTGGCCG	MID6 IGKJ5	atatcgcgagCTCCAGTCGTGTCCCTTGGCCG	MID10 IGKJ5	tgatacgtctCTCCAGTCGTGTCCCTTGGCCG
MID2 IGKC	acgctcgacaCCTTCCACTGTACTTTGGCCTC	MID6 IGKC	atatcgcgagCCTTCCACTGTACTTTGGCCTC	MID10 IGKC	tgatacgtctCCTTCCACTGTACTTTGGCCTC
MID3 IGKV1	agacgcactcCATCCAGWTGACCCAGTCTCC	MID7 IGKV1	cgtgtctctaCATCCAGWTGACCCAGTCTCC	MID11 IGKV1	tactgagctaCATCCAGWTGACCCAGTCTCC
MID3 IGKV2	agacgcactcGATATTGTGATGACCCAGWCT	MID7 IGKV2	cgtgtctctaGATATTGTGATGACCCAGWCT	MID11IGKV2	tactgagctaGATATTGTGATGACCCAGWCT
MID3 IGKV3	agacgcactcGACRCAGTCTCCAGCCACCCTG	MID7 IGKV3	cgtgtctctaGACRCAGTCTCCAGCCACCCTG	MID10 IGKV3	tactgagcta ACRCAGTCTCCAGCCACCCTG
MID3 IGKV4	agacgcactcGACATCGTGATGACCCAGTCT	MID7 IGKV4	cgtgtctctaGACATCGTGATGACCCAGTCT	MID11 IGKV4	tactgagctaACATCGTGATGACCCAGTCT
MID3 IGKV5	agacgcactcGAAACGACACTCACGCAGTCT	MID7 IGKV5	cgtgtctctaGAAACGACACTCACGCAGTCT	MID11 IGKV5	tactgagctaAAACGACACTCACGCAGTCT
MID3 IGKV6	agacgcactcGAAATTGTGCTGACTCAGTCT	MID7 IGKV6	cgtgtctctaGAAATTGTGCTGACTCAGTCT	MID11 IGKV6	tactgagctaGAAATTGTGCTGACTCAGTCT
MID3 IGKJ1/4	agacgcactcTCCACCTTGGTCCCTYSGCCG	MID7 IGKJ1/4	cgtgtctctaTCCACCTTGGTCCCTYSGCCG	MID11 IGKJ1/4	tactgagctaTCCACCTTGGTCCCTYSGCCG
MID3 IGKJ2	agacgcactcCTCCAGCTTGGTCCCCTGGCCA	MID7 IGKJ2	cgtgtctctaCTCCAGCTTGGTCCCCTGGCCA	MID11 IGKJ3	tactgagctaCTCCAGCTTGGTCCCCTGGCCA
MID3 IGKJ3	agacgcactcATCCACTTTGGTCCCAGGGCCG	MID7 IGKJ3	cgtgtctctaATCCACTTTGGTCCCAGGGCCG	MID11 IGKJ2	tactgagctaATCCACTTTGGTCCCAGGGCCG
MID3 IGKJ5	agacgcactcCTCCAGTCGTGTCCCTTGGCCG	MID7 IGKJ5	cgtgtctctaCTCCAGTCGTGTCCCTTGGCCG	MID11 IGKJ5	tactgagctaCTCCAGTCGTGTCCCTTGGCCG
MID3 IGKC	agacgcactcCCTTCCACTGTACTTTGGCCTC	MID7 IGKC	cgtgtctctaCCTTCCACTGTACTTTGGCCTC	MID11 IGKC	tactgagctaCCTTCCACTGTACTTTGGCCTC

Chapter 3

Lambda light chain gene rearrangements in mature naïve and IgA expressing B cells

3.1 Introduction

Previous studies from our laboratory have identified some unusual features of rearrangements of immunoglobulin IGL chain used by intestinal IgA plasma cells and IgA plasmablasts expressing high levels of β7 integrin in blood (Su, Boursier et al. 2004, Su, Gordon et al. 2008). It has been found that these cells are highly mutated as a result of SHM suggesting their development during the germinal centre reaction due to chronic antigenic challenge. In addition gene rearrangements involving *IGLV*1 and *IGLV*2 families, sampled from DNA, were significantly more often in-frame in the intestinal plasma as compared to frequencies in the published data of the naive mantle zone B cells (Farner, Dorner et al. 1999). However, out-of-frame gene rearrangements if present mainly involved *IGLV*5 family when studied in IgA single gut plasma cells or cell lines. Interestingly the out-of-frame gene rearrangements involved less frequent use of *IGLJ*1 gene segment in IgA plasma cells than mantle zone B cells. Variation in usage of *IGLJ* gene segments can be a reflection of light chain editing or revision. IGL revision has been considered as a mechanism for the diversification of the repertoire in intestinal IgA plasma cells by our group and this is supported by identification of more recombination excision circles (RECs) involving *IGLJ*2/3 than observed in naïve B cells (Su, Gordon et al. 2008). This chapter involves next generation sequencing to investigate the bias in IGL gene rearrangement in IgA expressing B cells.

Recently it has been reported that the T2 subset of transitional B cells show tropism towards GALT. These bone marrow emigrant transitional T2 B cells express integrin α4β7 and get activated in the GALT as shown by phosphorylation status of Syk, Btk and Erk. (Vossenkamper, Blair et al. 2013). The consequences of this gut tropism are not fully known, but one possibility is that transitional B cells may be the precursors of IgA plasma cells forming an independent pathway of development. This will be tested in this thesis by looking for features of *IGL* gene rearrangements considered to be characteristic of IgA plasma cells in transitional B cells.

3.1.1 Aims of this chapter

1. Is the bias towards in-frame gene rearrangements involving *IGLV*1 and *IGLV*2 gene families as observed previously in IgA secreting cells also evident in transitional B cells?
2. What is the distribution of in-frame and out-of-frame *IGL* gene rearrangement in mature naïve and IgA expressing B cell subsets obtained by high throughput sequencing?
3. Are there any differences in the inherent biases during gene rearrangement at the *IGL* locus between mature and naïve and IgA expressing B cells?
4. Is there any preferential selection of *IGLV* gene segment during the IgA response?
5. Is there any evidence of secondary gene rearrangements in mature naïve and IgA expressing B cells detectable by analysis of *IGLJ* gene segments?
6. Is there any evidence of a bias in light chain expression in IgA by flow cytometry?

3.2 Results

3.2.1 Is the bias towards in-frame gene rearrangements in *IGLV*1 and *IGLV*2 genes observed previously in IgA secreting cells observed in transitional B cells?

In previous studies mucosal IgA plasma cells and plasmablasts have *IGL* gene rearrangements that were biased towards in-frame rearrangements involving *IGLV*1 and *IGLV*2 gene families compared to the published studies of blood cells or naïve B cells in gut, when these were sampled in DNA by family specific PCR reactions. The first aim of this thesis was to determine if the excess of in-frame gene rearrangements observed in *IGLV*1 and *IGLV*2 family in mucosal IgA expressing B cells but not naïve B cells is also observed in the transitional B cells.

PBMCs from three healthy anonymous blood donors were sorted into four B cell subsets: transitional, mature naïve, IgA plasmablasts and IgA memory B cells. The gating strategy is shown in Figure 3-1 and 3-2. The DNA was extracted and gene families *IGLV1* and *IGLV2* were amplified by family specific PCR and products were cloned in pGEM-T vector after purification. After screening clones for insert of the correct size, selected clones were sequenced and gene rearrangements with unique CDR-3 sequences were analyzed.

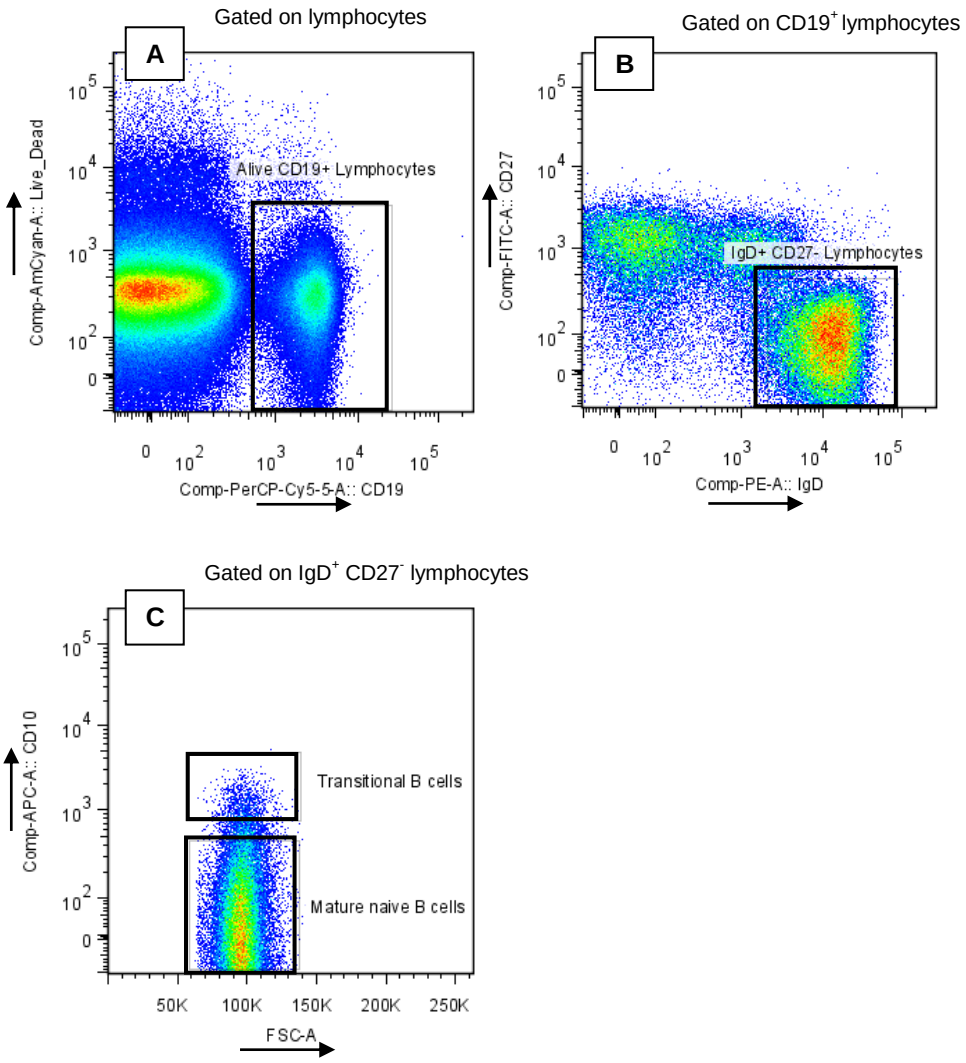


Figure 3-1: Sorting of B cells into transitional and mature naïve B cell subsets. Alive B cells **(A)** were gated and classified to the $IgD^{+}CD27^{-}$ B cells **(B)**. These cells were divided according to the expression of CD10 into transitional ($CD19^{+}IgD^{+}CD27^{-}CD10^{hi}$) and mature naïve ($CD19^{+}IgD^{+}CD27^{-}$ B cells) B cells **(C)**.

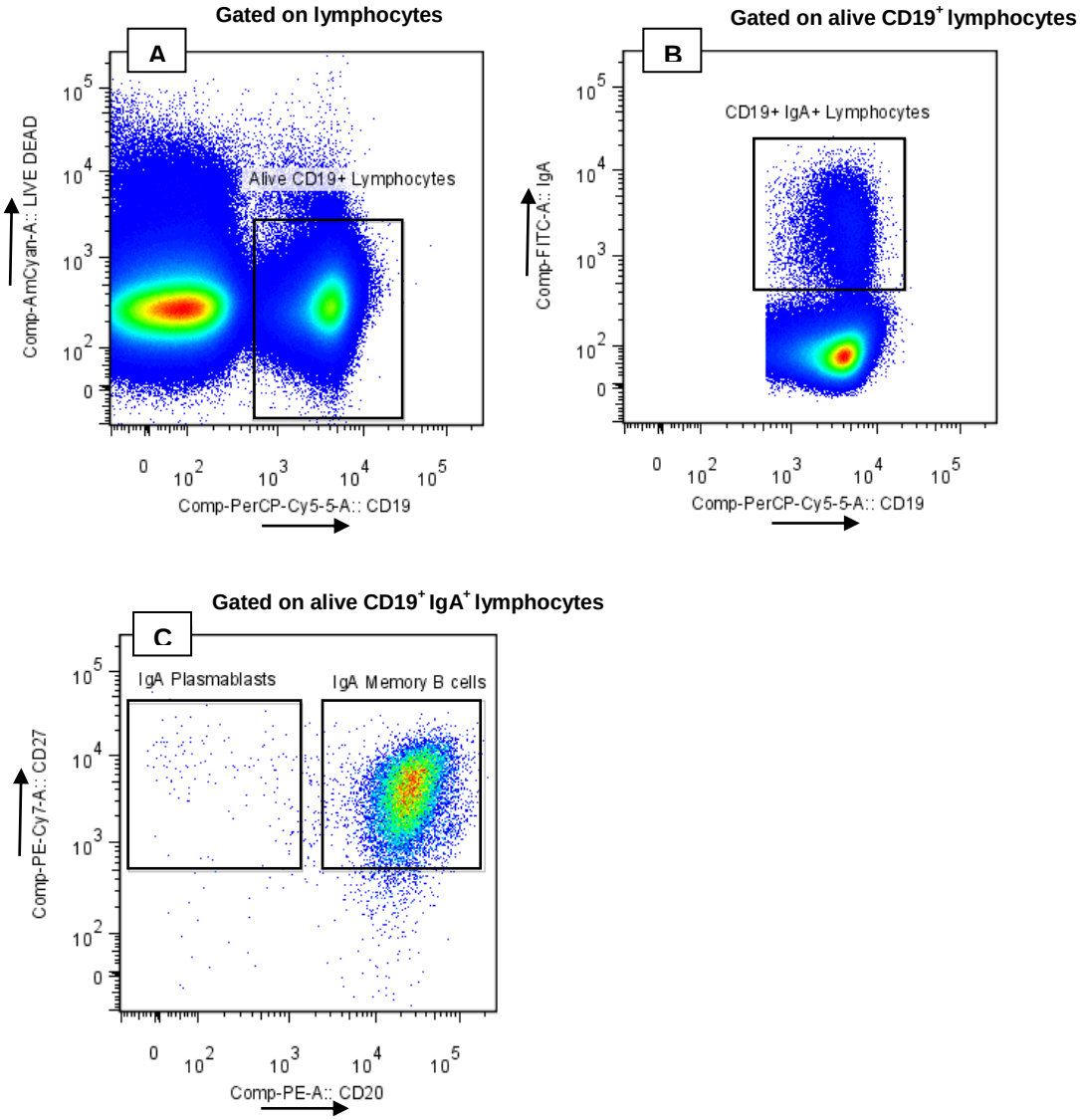


Figure 3-2: Sorting of B cells into IgA plasmablasts and IgA memory B cells. Alive B cells **(A)** were classified into IgA expressing lymphocytes **(B)**. These cells were divided according to the expression of CD20 and CD27 into IgA plasmablasts ($CD19^+CD20^-CD27^+IgA^+$) and IgA memory ($CD19^+CD20^+CD27^+IgA^+$) B cells **(C)**.

The number of unique DNA gene rearrangements obtained after sequencing are given in Table 3-1. Only 10 unique sequences were obtained from IgA plasmablasts in this study therefore, sequences from IgA plasmablasts were included from a previous study conducted in our lab and used for the comparison. It was found that the ratios of in-frame to out-of-frame gene rearrangements of transitional and naïve B cells appeared similar to naïve B cells in previous studies. The transitional and naïve B cells showed approximately 70% in-frame gene rearrangements. However IgA expressing B cells had more than 85% in-frame gene rearrangements (Figure 3-3). Therefore no resemblance between the features of transitional B cells and IgA plasma cells was identified by this analysis.

Table 3-1: Unique DNA gene rearrangements (*IGLV*1/2-*IGLJ*) derived after cloning

B cell subsets	Unique DNA gene rearrangements
Transitional B cells	29
Mature naïve B cells	31
IgA memory B cells	27
IgA plasmablasts (New)	10
IgA plasmablasts (Previous)	30

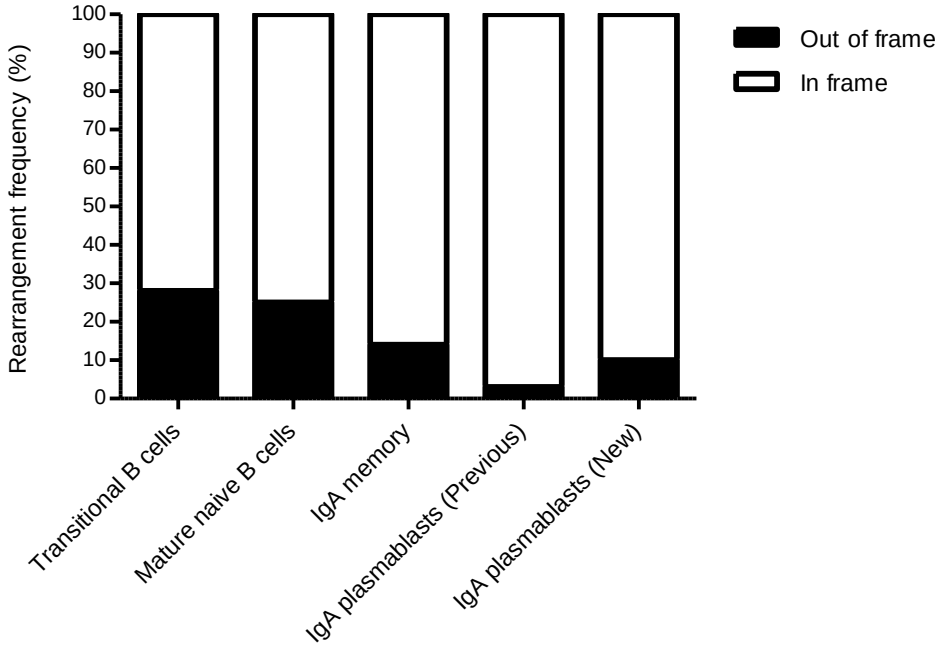


Figure 3-3: The ratios of in-frame to out-of-frame gene *IGLV*1/2 gene rearrangements in various B cell subsets. Transitional, naïve, IgA plasmablasts and IgA memory B cells were isolated from healthy blood donors. *IGL* genes (*IGLV*1/2 – *IGLJ*) were amplified, purified, cloned and sequenced. Transitional and naïve B cells have more out-of-frame gene rearrangements than IgA expressing B cells. Sequences of IgA plasmablasts (Previous) were used from the previous study.

Analysis of rearrangements of IGL from mature naïve and IgA expressing B cells by next generation sequencing

Two major problems were encountered in the previous section. Firstly the number of different sequences obtained was very low. Secondly it is difficult to understand the significance of rearrangements of *IGLV*1 and *IGLV*2 families when the whole repertoire is not included in the analysis. Therefore, in order to get better insight a high throughput sequencing approach was used, amplifying all *IGL* families simultaneously. Furthermore, in this experiment, IgA expressing B cell subsets in blood were classified further into IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$IgA cells.

The mature naïve and three subsets of IgA expressing B cells from three healthy individuals were sorted into separate tubes containing RPMI medium supplemented with FCS and antibiotics (streptomycin and penicillin)(Table 3-2). The gating strategy plots for identifying various B cell subsets are shown in Figures 3-1 and 3-4.

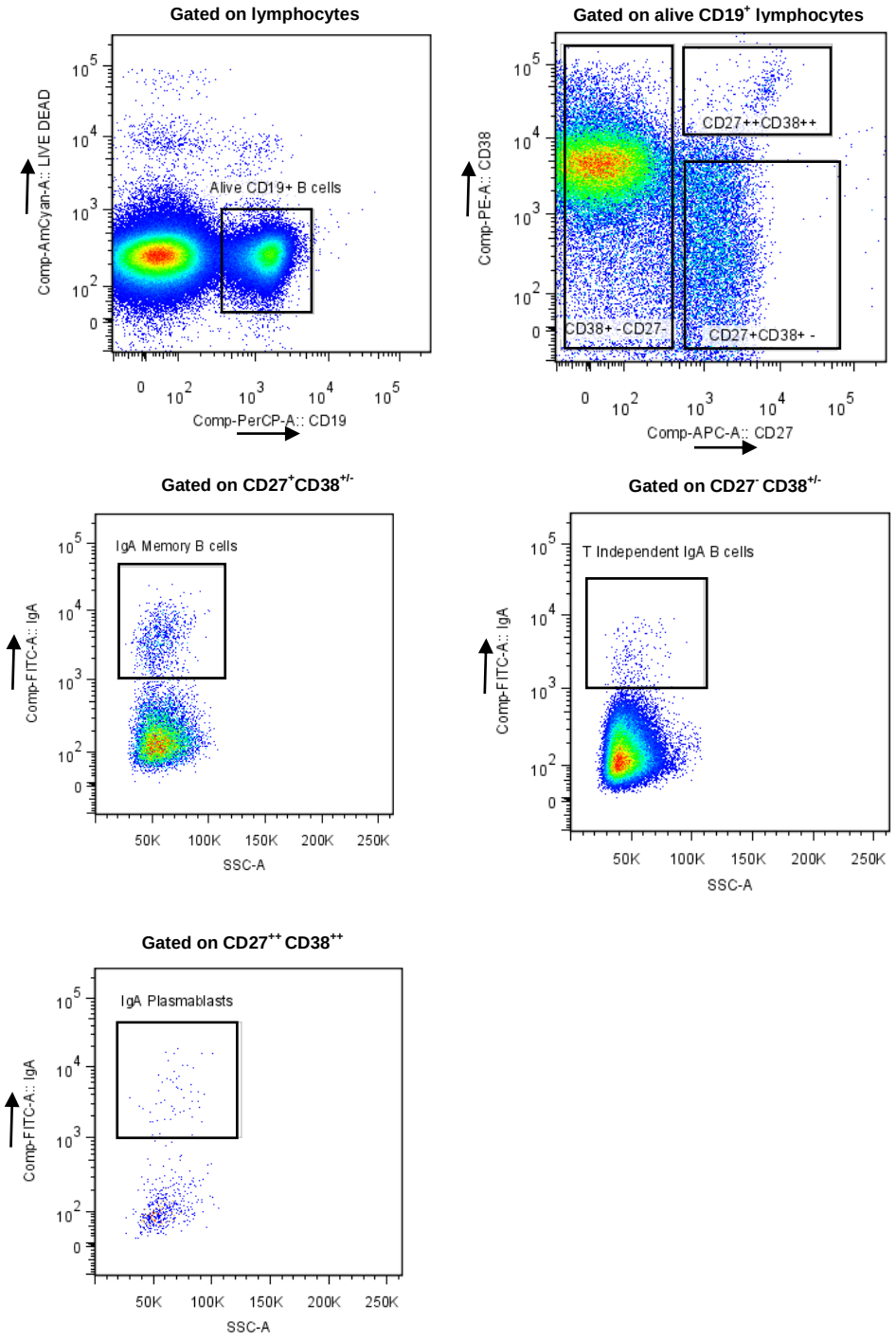


Figure 3-4: Sorting of B cells into IgA expressing B cells. Alive $CD19^+$ B cells were classified into $CD27^-CD38^{+/-}$, $CD27^+CD38^{+/-}$ and $CD27^{++}CD38^{++}$. Furthermore, based on surface IgA division was made into $CD27^-CD38^{+/-}IgA^+$, $CD27^+CD38^{+/-}IgA^+$ and IgA plasmablasts ($CD27^{++}CD38^{++}$ IgA^+).

Table 3-2: The respective number of mature naïve and subsets of IgA expressing B cells derived from three healthy donors (HD)

Sample	B cell subset	Number of cells
HD-1	Mature naïve B cells	115,023
	$CD27^{+}CD38^{+/-}$IgA	21,705
	$CD27^{-}CD38^{+/-}$IgA	29,100
	$CD27^{++}CD38^{++}$IgA	8,396
HD-2	Mature naïve B cells	90,199
	$CD27^{+}CD38^{+/-}$IgA	33,236
	$CD27^{-}CD38^{+/-}$IgA	12,966
	$CD27^{++}CD38^{++}$IgA	2,210
HD-3	Mature naïve B cells	12,1080
	$CD27^{+}CD38^{+/-}$IgA	46,728
	$CD27^{-}CD38^{+/-}$IgA	11,332
	$CD27^{++}CD38^{++}$IgA	22,436

The DNA was extracted from all B cell subsets. For the amplification of rearranged *IGL* repertoire multiplex primers from *IGLV* to *IGLJ* were used. In order to distinguish various amplified products for the second round of PCR, primers tagged with different bar codes consisting of 10 nucleotides at the end called MID were used. Various tags used for the differentiation are given in Table 3-3.

Table 3-3: The primer tags used for differentiating the amplified products from three healthy donors

Sample	B cell subset	MID tag
HD-1	Mature naïve B cells	MID-1
	$CD27^{+}CD38^{+/-}IgA$	MID-2
	$CD27^{-}CD38^{+/-}IgA$	MID-3
	$CD27^{++}CD38^{++}IgA$	MID-4
HD-2	Mature naïve B cells	MID-5
	$CD27^{+}CD38^{+/-}IgA$	MID-6
	$CD27^{-}CD38^{+/-}IgA$	MID-7
	$CD27^{++}CD38^{++}IgA$	MID-8
HD-3	Mature naïve B cells	MID-9
	$CD27^{+}CD38^{+/-}IgA$	MID-10
	$CD27^{-}CD38^{+/-}IgA$	MID-11
	$CD27^{++}CD38^{++}IgA$	MID-12

The amplified products of appropriate sizes (350 bp) were excised from the agarose gel and purified and quantified using Qubit flurorometer. Samples were sequenced using GS FLX Genome Titanium Sequencer 454 and resulting sequences were passed through quality control criteria (as described in section 2.3.6). Numbers of unique DNA gene rearrangements obtained from different B cell subsets are given in Table 3-4.

Table 3-4: The respective number of unique DNA gene rearrangements derived from mature naïve and IgA expressing B cell subsets from three healthy donors

B cell subset	Unique *IGLV-J* gene rearrangements
Mature naïve B cells	1435
IgA expressing B cells	1480
1. $CD27^{+}CD38^{+/-}$IgA	1123
2. $CD27^{-}CD38^{+/-}$IgA	234
3. $CD27^{++}CD38^{++}$IgA	123

3.2.2 What is the distribution of in-frame and out-of-frame *IGL* gene rearrangement in mature naïve and IgA expressing B cell subsets obtained by high throughput sequencing?

Gene rearrangements were classified as in-frame and out-of-frame according to the functionality of the junction after high throughput sequencing designated by IMGT. It was found that there were 55% in-frame gene rearrangements in mature naïve B cells. However, in IgA expressing B cells 81% gene rearrangements were in the correct genetic reading frame (Figure 3-5).

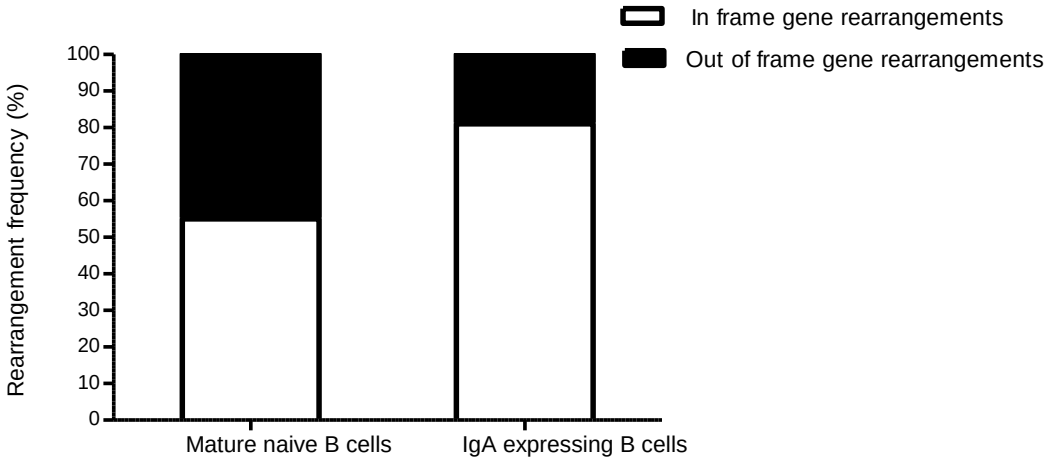


Figure 3-5: The relative distribution of in-frame and out-of-frame gene rearrangements in mature naïve and IgA expressing B cells. Gene rearrangements were classified according to the functionality of CDR-3 junction. There were relatively more in-frame gene rearrangements in IgA expressing B cells than naïve B cells.

In order to determine if this relative abundance of in-frame gene rearrangements is evident in different subsets of IgA expressing B cells the relative distribution was determined in each subset. It was found that as compared to naïve B cells there tended to be more in-frame gene rearrangements in IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$IgA as shown in Figure 3-6.

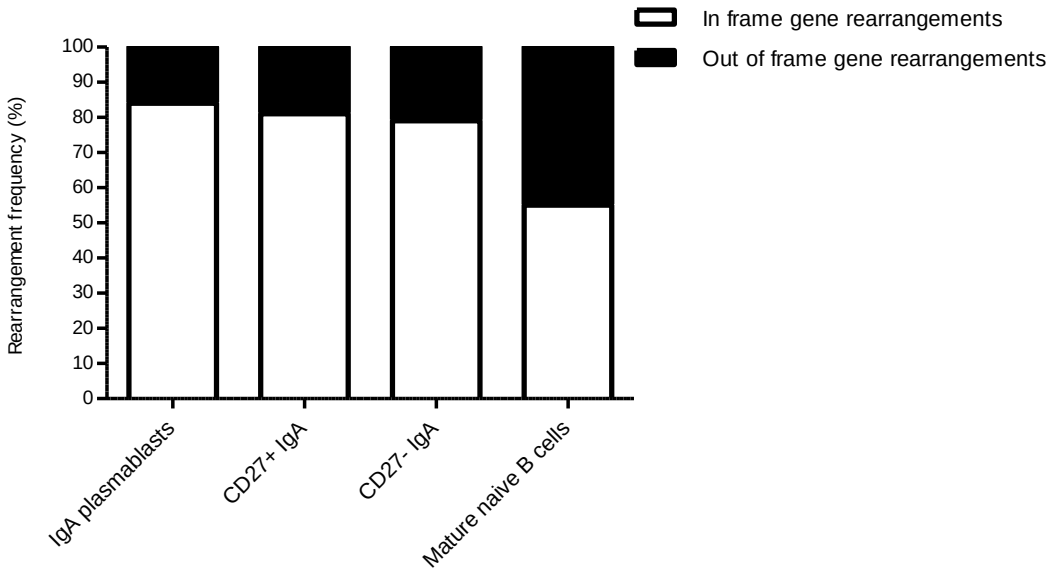


Figure 3-6: The relative distribution of in-frame and out-of-frame gene rearrangements in various subsets of IgA expressing B cells. IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$IgA exhibited overall 80% in frame gene rearrangements.

Since there was no apparent difference in the subsets of IgA expressing cells in the ratios of in-frame and out-of-frame gene rearrangements they were pooled in the next series of analyses. It was found that *IGLV*5 and *IGLV*9 families comprised more out-of-frame gene than in-frame gene rearrangements in both mature naïve B cells and IgA expressing cells (Figure 3-7 and 3-8). Comparing data in Figures 3-8 and 3-9 it appears that the *IGLV* 1, 2 and 3 families have a greater abundance of out-of-frame gene rearrangements in naïve as compared to IgA expressing B cells as in previous data.

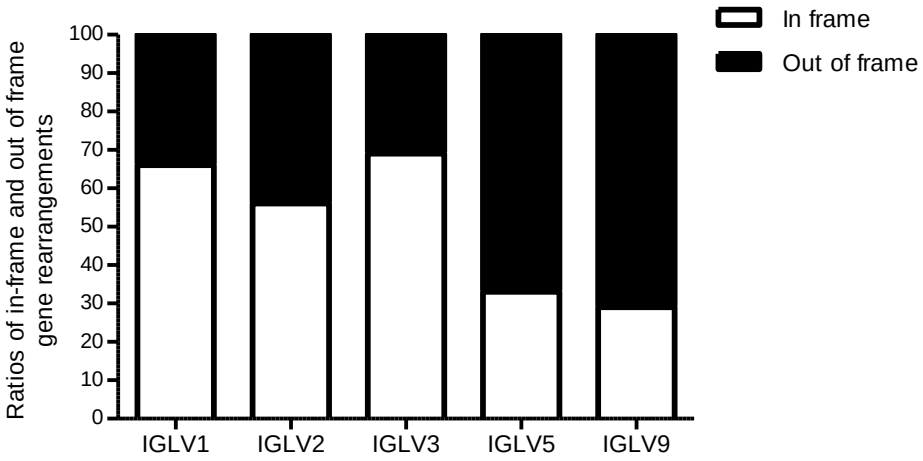


Figure 3-7: Ratios of in-frame to out-of-frame gene rearrangements in mature naïve B cells. Mature naïve B cells were isolated from three healthy donors. *IGL* genes were amplified, purified, deep sequenced and classified as in-frame or out-of-frame. Gene families *IGLV*5 and *IGLV*9 had more out-of-frame than in-frame gene rearrangements.

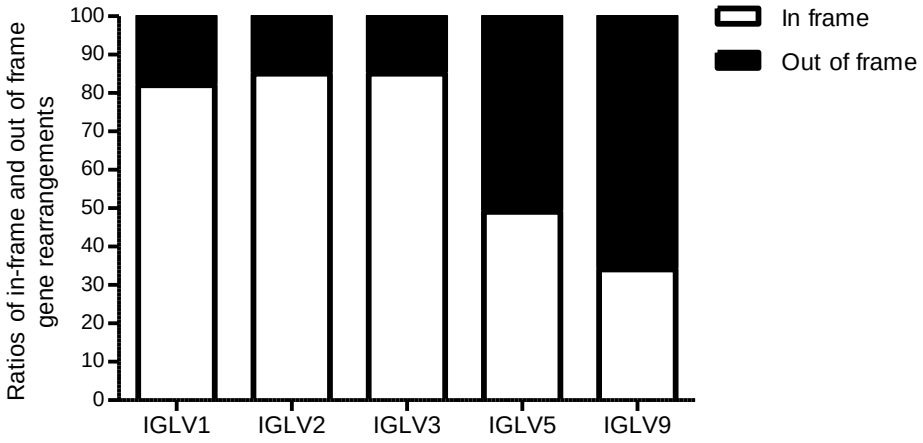


Figure 3-8: Ratios of in-frame to out-of-frame gene rearrangements in IgA expressing B cells. IgA expressing B cells were isolated from three healthy donors. *IGL* genes were amplified, purified, deep sequenced and classified as in-frame or out-of-frame. Gene families *IGLV*5 and *IGLV*9 had more out-of-frame than in-frame gene rearrangements.

The data above was expressed as relative frequencies of in-frame and out-of-frame gene rearrangements within families because previous data generated by family specific PCR had been expressed in this way. In Figure 3-9 the relative distribution of different functionalities (of the CDR-3 junction) out of total gene rearrangements has been shown in mature naïve and IgA expressing cells separately. It was found that gene family *IGLV*2 was abundantly rearranged in both subsets. However unlike mature naïve B cells, IgA expressing B cells appeared to have more in-frame and less out-of-frame gene rearrangements of *IGLV*2 family and this resulted in relative reduction of other gene families analyzed.

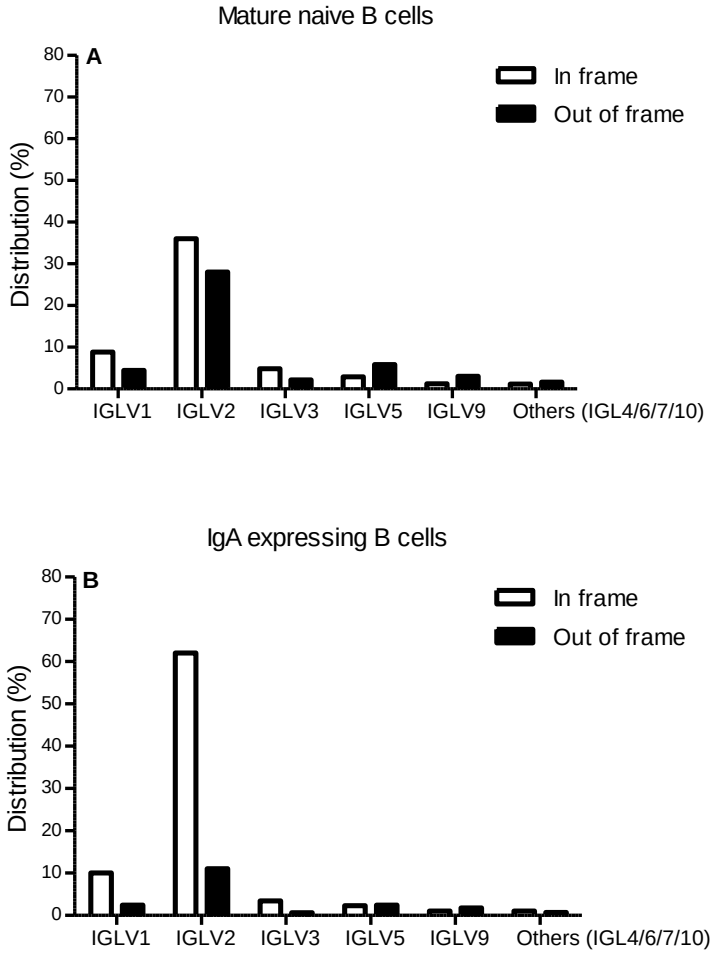


Figure 3-9: Distribution of in-frame and out-of-frame gene rearrangements out of total gene rearrangements obtained from the respective mature naïve (A) and IgA expressing B cells (B). Gene family *IGLV*2 was found to be frequently rearranged. However, there was a reduction in the out-of-frame gene rearrangements in IgA expressing B cells.

The next step was to determine, if the relative abundance of in-frame gene rearrangements involving *IGLV*2 family in IgA expressing B cells was evident during rearrangement with all *IGLJ* families. For this purpose, the distribution of the functionalities of CDR-3 junction was determined in gene rearrangements with *IGLJ*1 and *IGLJ*2/3 separately. A similar trend was observed during rearrangement with either *IGLJ*1 or *IGLJ*2/3 irrespective of B cell subsets. The gene family *IGLV*2 was frequently rearranged with both *IGLJ*1 and *IGLJ*2/3 with a marked dominance of in-frame gene rearrangements in the *IGLV*2 family (Figure 3-10 and 3-11).

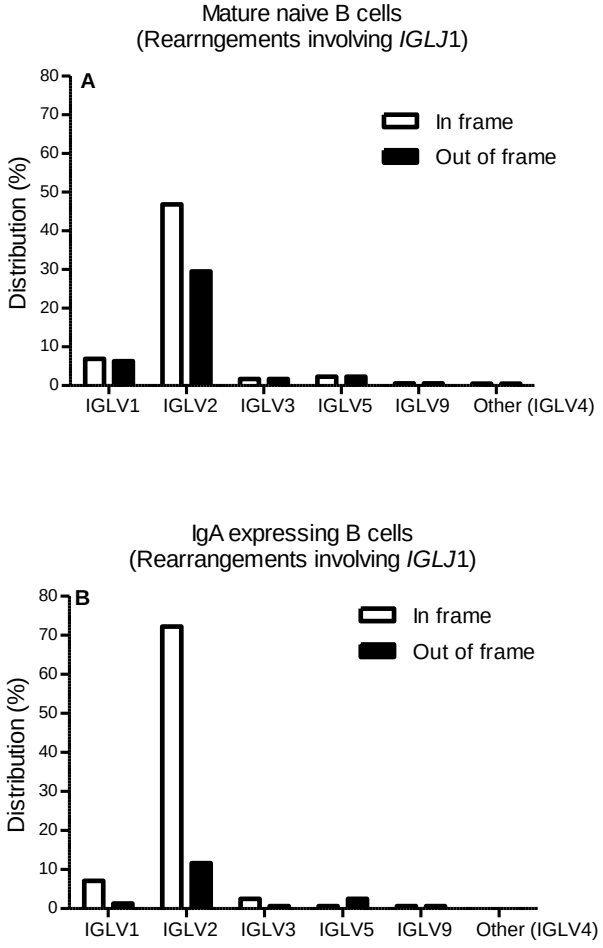


Figure 3-10: Distribution of in-frame and out-of-frame gene rearrangements of different *IGLV* families with *IGLJ*1 out of total gene rearrangements obtained from respective mature naïve (A) and IgA expressing (B) B cells. Gene family *IGLV*2 was found to be frequently rearranged. However, there was a reduction in the out-of-frame gene rearrangements in IgA expressing B cells.

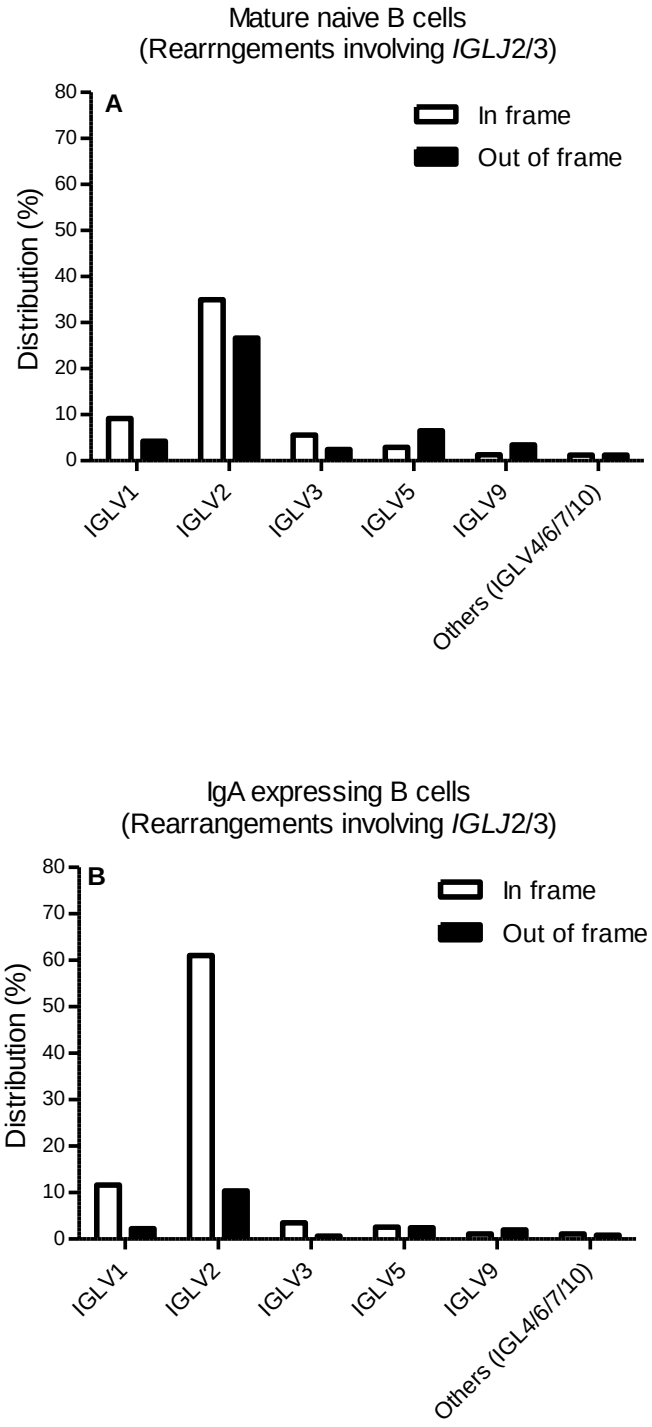


Figure 3-11: Distribution of in-frame and out-of-frame gene rearrangements of different *IGLV* families with *IGLJ2/3* out of total gene rearrangements obtained from respective mature naïve (A) and IgA expressing (B) B cells. Gene family *IGLV*2 was found to be frequently rearranged. However, there was a reduction in the out-of-frame gene rearrangements in IgA expressing B cells.

The proportions of in-frame:out-of-frame gene rearrangements in *IGLV*5 and 9 families in Figures 3-7 and 3-8 resembled the 1:2 expected ratio of IF:OF if rearrangements were all non functional so that the ratio would not be affected by selection into the used pool of sequences. To test if *IGLV*5 and 9 gene families are used, usage of *IGL* segments was determined by retrieving published data (references are in figure's legend). As shown in Figure 3-12 none of 210 sequences used *IGLV*5 or 9 suggesting that despite abundant productive gene rearrangements *IGLV*5 and 9 are either not used or rarely used whether in-frame or out-of-frame.

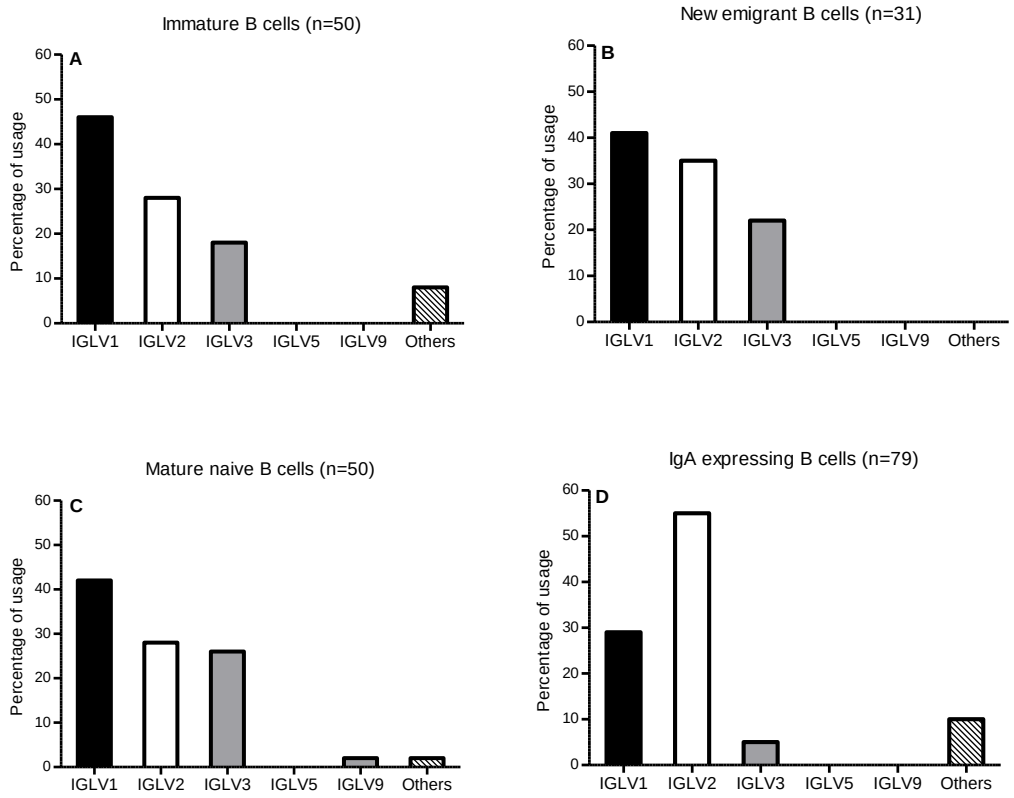


Figure 3-12: Distribution of various *IGL* families in the expressed repertoire (cDNA) of various B cell subsets (A-D). *IGLV*5 and 9 were absent in the expressed gene rearrangements and *IGLV*2 is the most widely used family in IgA expressing B cells (Wardemann, Yurasov et al. 2003, Yurasov, Tiller et al. 2006, Benckert, Schmolka et al. 2011, Steinsbo, Henry Dunand et al. 2014).

3.2.3 Are there any differences in the inherent biases during gene rearrangement at the *IGL* locus between mature naïve and IgA expressing B cells?

Six hundred and forty five out-of-frame gene rearrangements from naïve B cells were compared with 286 out-of-frame gene rearrangements from IgA expressing B cells. No difference was observed in the relative abundance of different *IGLV* families (Figure 3-13). When usage of individual *IGLV* gene segments was compared, again there was no difference in the relative rearrangement frequencies of various *IGLV* gene segments between different B cell subsets (Figure 3-14). Overall, the cluster A (comprised of *IGLV2* and 3 families) was abundantly rearranged followed by the cluster B and C. In terms of gene segments, *IGLV2-14*, *IGLV2-23*, *IGLV2-8*, *IGLV2-5*, *IGLV9-49* and *IGLV1-47* were frequently rearranged in both mature naïve and IgA expressing B cells.

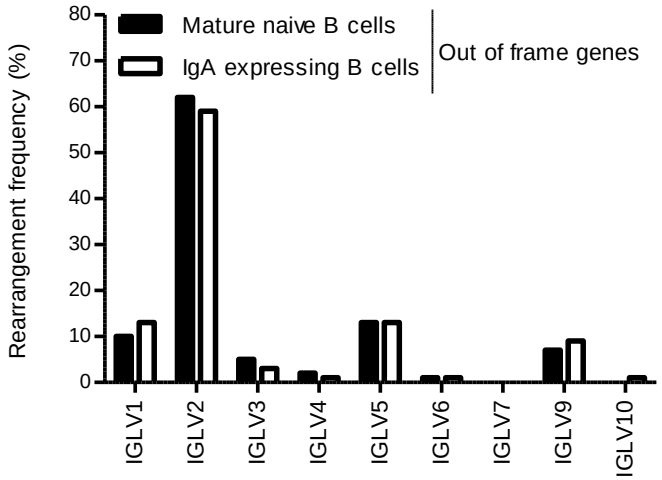


Figure 3-13: Distribution of *IGLV* gene families in the out-of-frame gene rearrangements. Comparison of relative rearrangement frequencies of *IGLV* gene families between out-of-frame gene rearrangements obtained from mature naïve and IgA expressing B cells. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). There was no difference in the relative distribution of *IGLV* gene families between mature naïve and IgA expressing B cells.

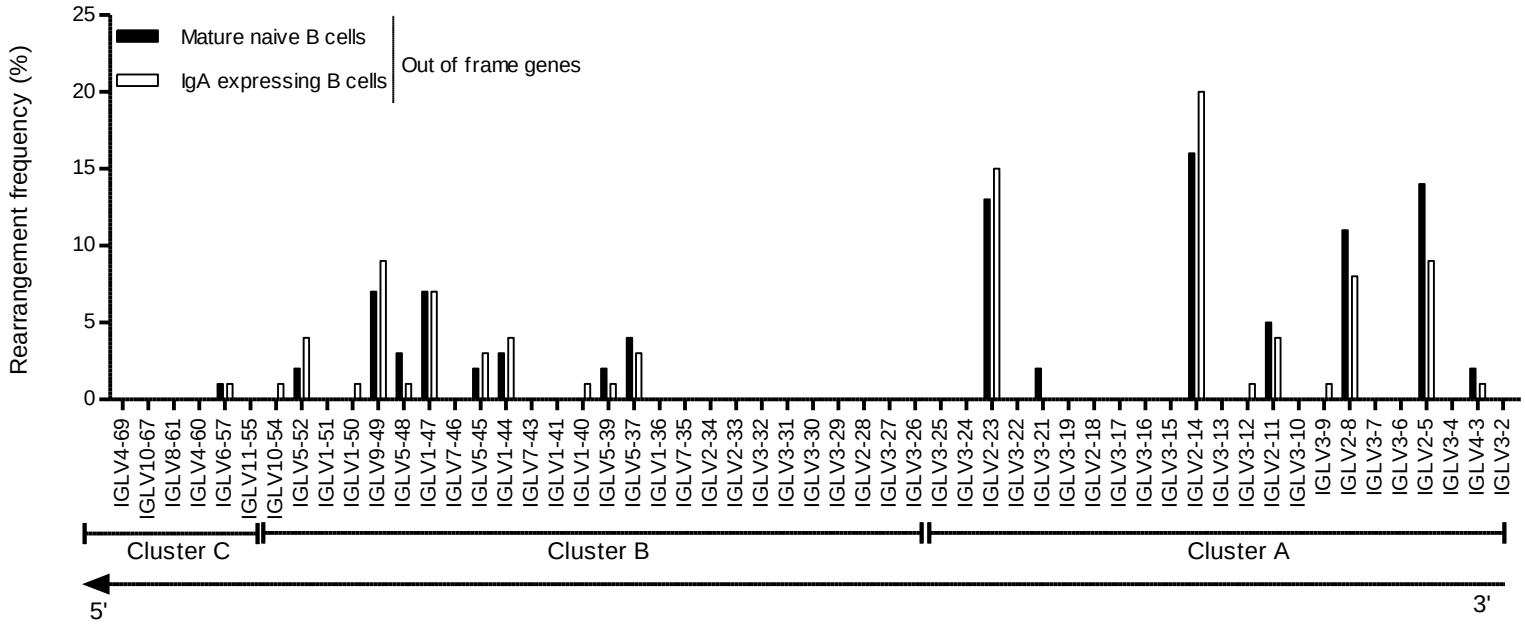


Figure 3-14: Analysis of out-of-frame *IGLV* gene segments. Comparison of relative rearrangement frequencies of out-of-frame *IGLV* gene segments between mature naïve and IgA expressing B cells isolated from three healthy individuals (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). There was no difference in the relative distribution of *IGLV* gene segments. The gene segments are arranged in order of their position on chromosome 22q11.2.

3.2.4 Are there any *IGLV* gene segments preferentially selected during the IgA response?

The relative frequencies of in-frame gene rearrangements of *IGLV* families in naïve and IgA expressing B cells were compared. The *IGLV*2 gene family was significantly more common in IgA expressing B cells than naïve B cells while *IGLV*3 and *IGLV*5 gene families were comparatively more in mature naïve B cells (Figure 3-15).

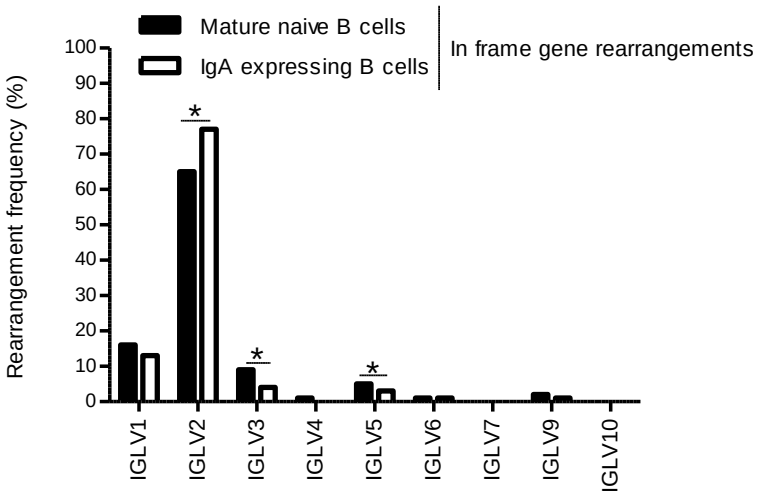
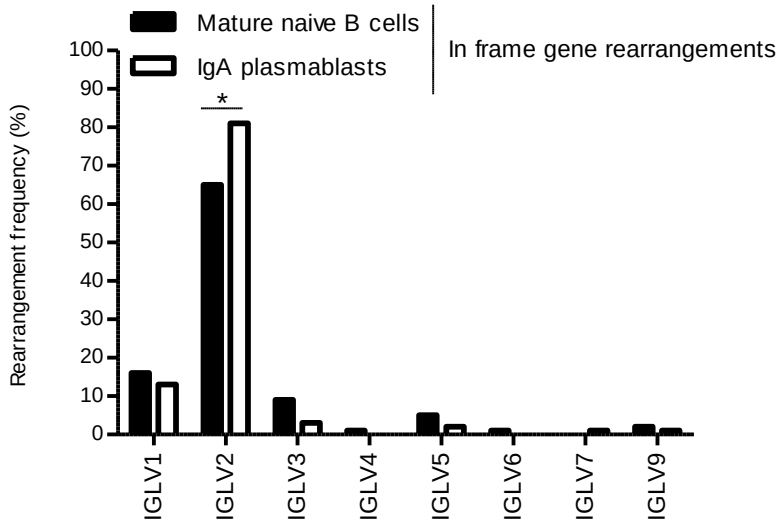
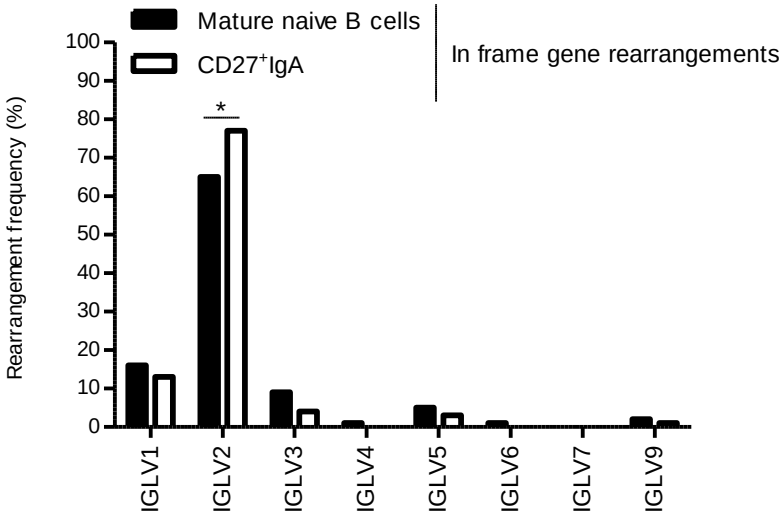


Figure 3-15: Distribution of *IGLV* gene families in the in-frame gene rearrangements. Comparison of relative rearrangement frequencies of *IGLV* gene families between in-frame gene rearrangements obtained from mature naïve and IgA expressing B cells. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). *IGLV2* gene family was significantly used more in IgA expressing B cells and gene families *IGLV3* and *IGLV5* were represented more in mature naïve B cells.

To investigate if the *IGLV2* gene family is more abundant in all IgA subsets, sequences from IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$ IgA were analyzed separately. It was found that the *IGLV2* gene family was more commonly in-frame in all subsets of IgA as compared to naïve B cells (Figure 3-16).


Mature naive B cells
IgA plasmablasts
In frame gene rearrangements
Rearrangement frequency (%)
100
90
80
70
60
50
40
30
20
10
0
*
IGLV1
IGLV2
IGLV3
IGLV4
IGLV5
IGLV6
IGLV7
IGLV9


Mature naive B cells
CD27+IgA
In frame gene rearrangements
Rearrangement frequency (%)
100
90
80
70
60
50
40
30
20
10
0
*
IGLV1
IGLV2
IGLV3
IGLV4
IGLV5
IGLV6
IGLV7
IGLV9

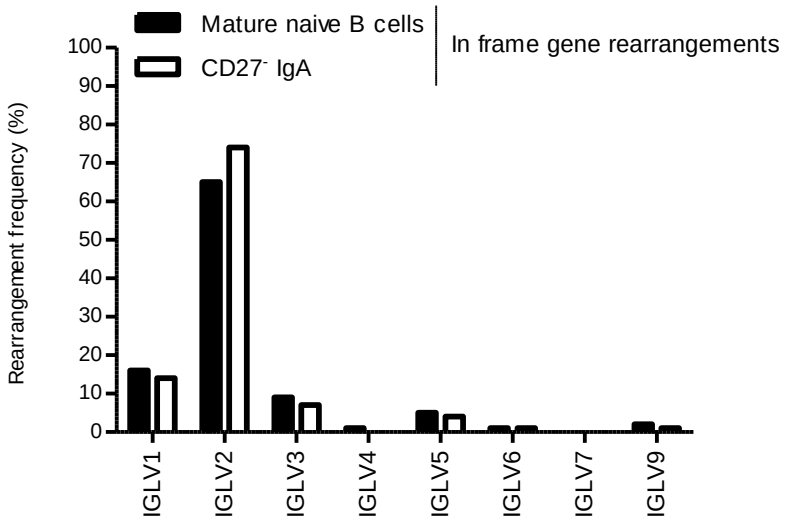


Figure 3-16: Distribution of *IGLV* gene families in the in-frame gene rearrangements. Comparison of relative rearrangement frequencies of *IGLV* gene families between in-frame gene rearrangements obtained from mature naïve and IgA expressing B cells (IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$IgA). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene family *IGLV2* was significantly used more in IgA expressing B cells.

Further investigation was done by comparing the relative usage of *IGLV* gene segments. It was found *IGLV2-14* was significantly more common in IgA expressing B cells suggesting preferential selection to use this segment. In contrast, *IGLV2-5* and *IGLV3-1* were significantly more abundant in the gene rearrangements of mature naïve B cells (Figure 3-17).

The relative distribution of *IGLV* gene segments was further explored in IgA subsets separately. It was found in IgA plasmablasts, $CD27^{+}$IgA and $CD27^{-}$IgA subset there was preferential usage of gene segment *IGLV2-14* though this was not statistically significant in the $CD27^{-}$IgA subset. As above, gene segments *IGLV2-5* and *IGLV3-1* tended to be relatively more abundant in mature naïve B cells than in any of the other B cell subsets (Figure 3-18 to 3-20).

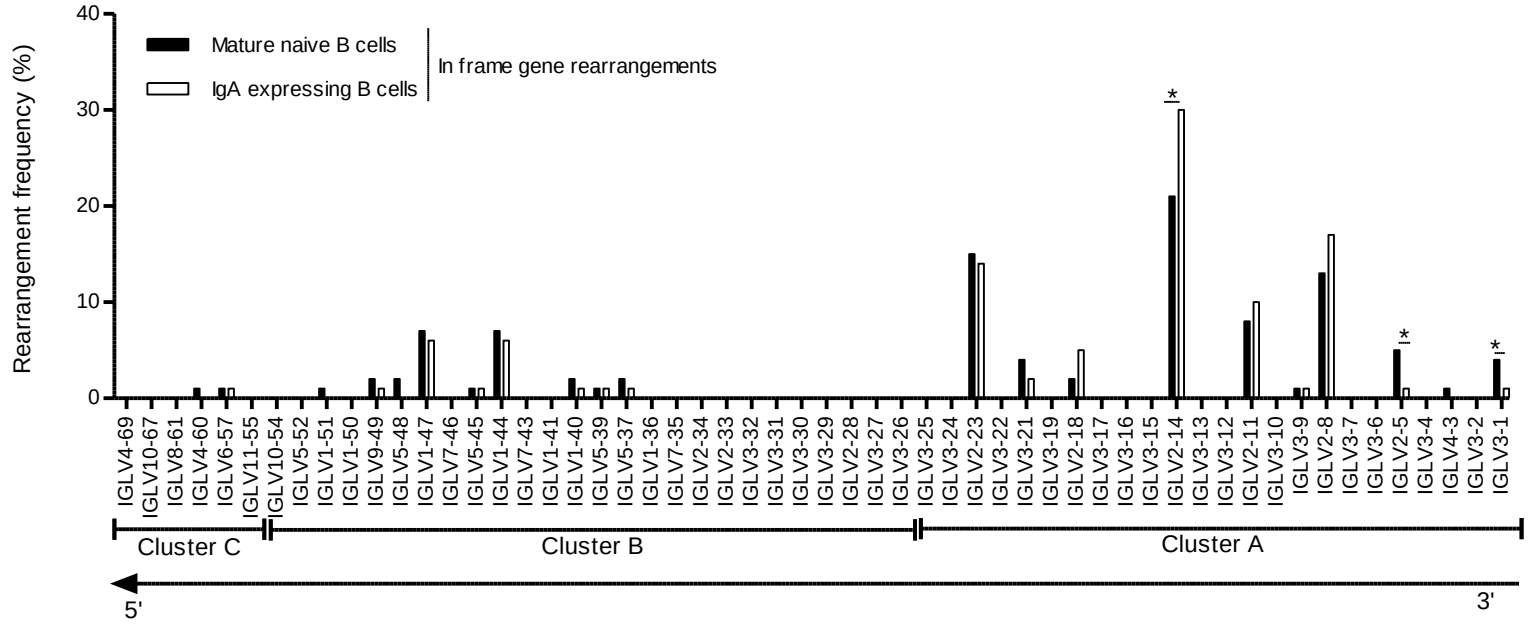


Figure 3-17: **Analysis of in-frame *IGLV* gene segments. Comparison of relative rearrangement frequencies of in-frame *IGLV* gene segments between mature naïve and IgA expressing B cells isolated from three healthy individuals (HD-1 to HD-3).** Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGLV2-14* was significantly higher in IgA expressing B cells while *IGLV2*-5 and *IGLV3-1* were significantly more in mature naïve B cells. The gene segments are arranged in order of their position on chromosome 22q11.2.

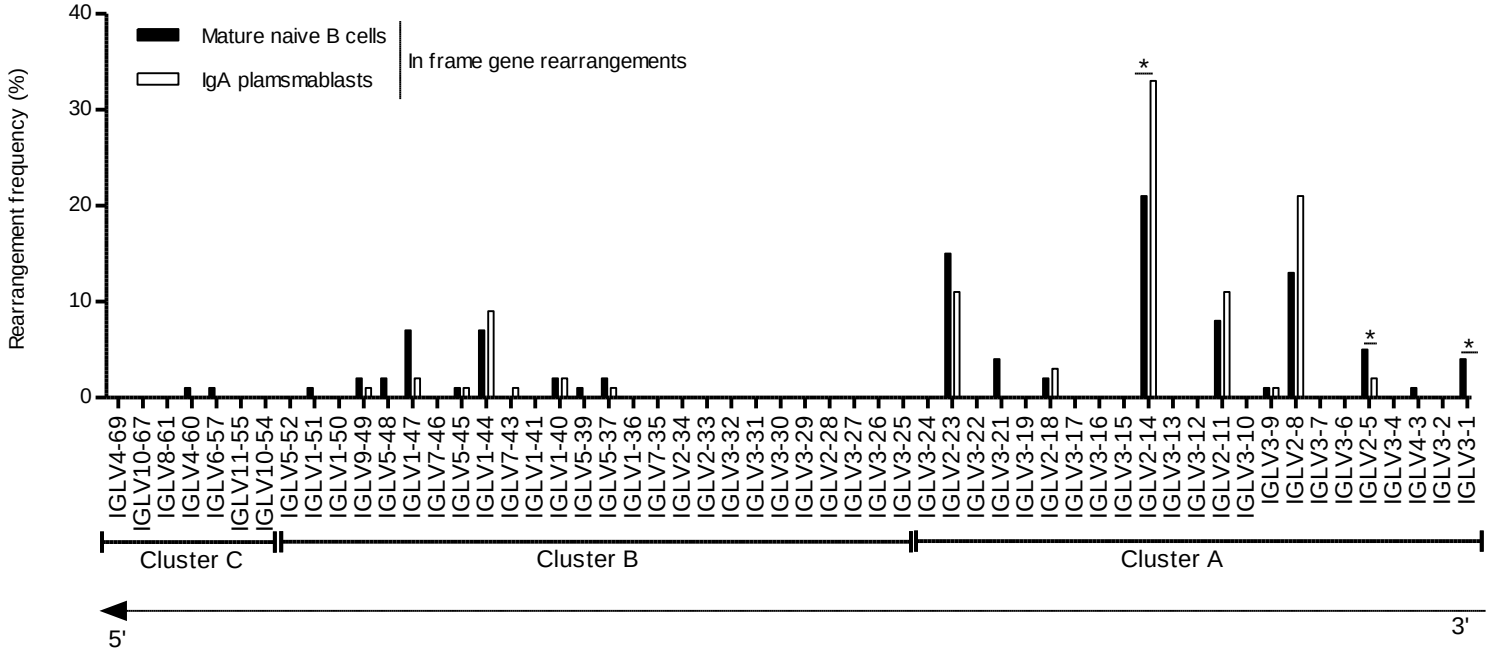


Figure 3-18: Analysis of in-frame *IGLV* gene segments. Comparison of relative rearrangement frequencies of in-frame *IGLV* gene segments between mature naïve and IgA plasmablasts isolated from three healthy individuals (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGLV2-14* was significantly higher in IgA plasmablasts while *IGLV2*-5 and *IGLV3-1* were significantly more in mature naïve B cells. The gene segments are arranged in order of their position on chromosome 22q11.2.

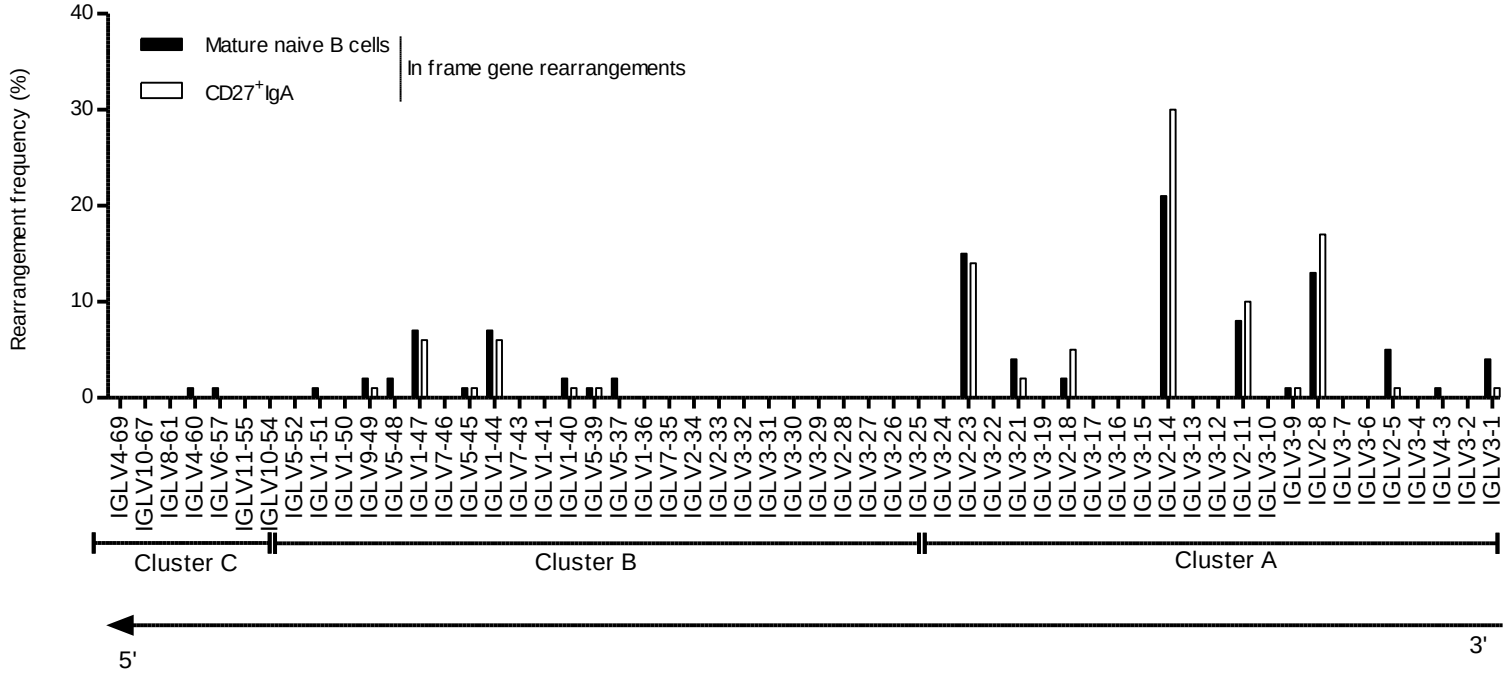


Figure 3-19: Analysis of in-frame *IGLV* gene segments. Comparison of relative rearrangement frequencies of in-frame *IGLV* gene segments between mature naïve and CD27$^+$IgA B cells isolated from three healthy individuals (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGLV2-14* was preferentially more in IgA memory B cells while *IGLV2*-5 and *IGLV3-1* were more in mature naïve B cells. The gene segments are arranged in order of their position on chromosome 22q11.2.

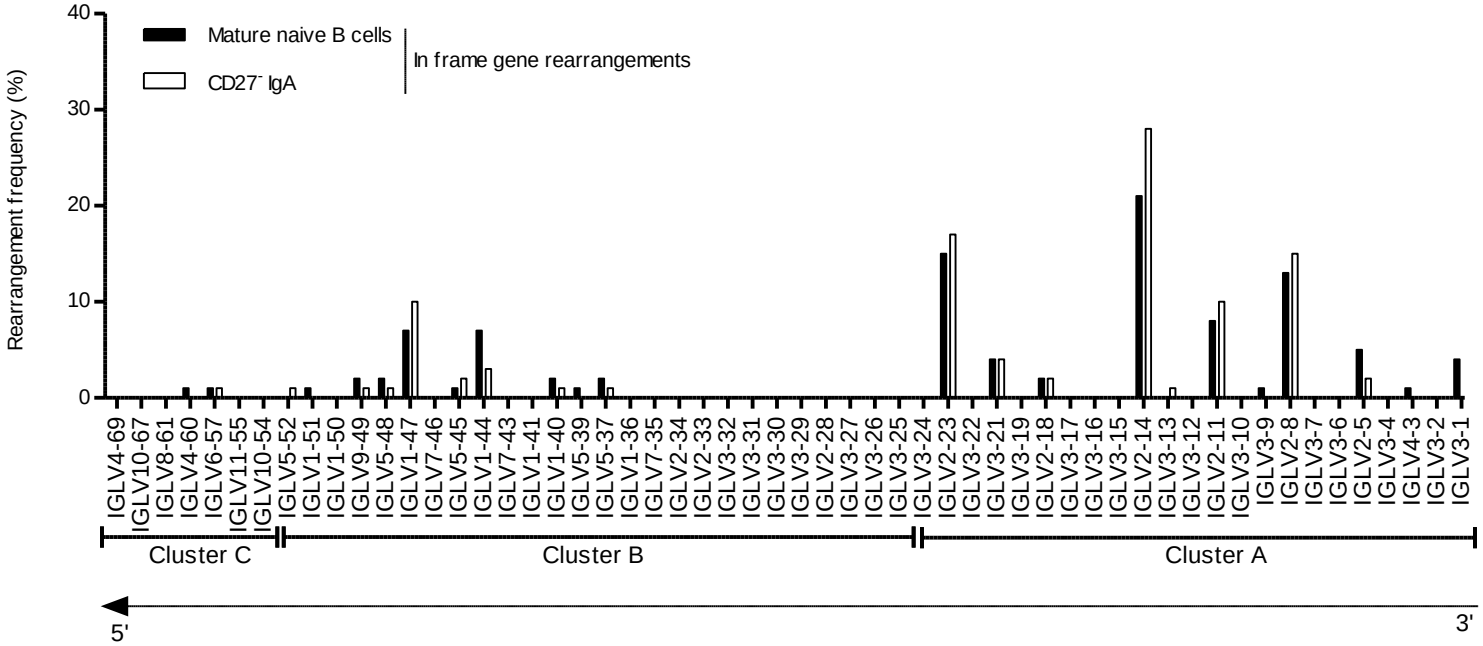


Figure 3-20: Analysis of in-frame *IGLV* gene segments. Comparison of relative rearrangement frequencies of in-frame *IGLV* gene segments between mature naïve and CD27$^-$IgA cells isolated from three healthy individuals (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGLV2-14* was preferentially more in CD27$^-$ IgA while *IGLV2*-5 and *IGLV3-1* were more in mature naïve B cells. The gene segments are arranged in order of their position on chromosome 22q11.2

3.2.5 Is there any evidence of secondary gene rearrangements in mature naïve and IgA expressing B cells detectable by analysis of *IGLJ* gene segments?

Rearranged *IGL* chain loci may retain unrearranged 5' *IGLV* and 3' *IGLJ* gene segments with correctly oriented RSS and thus providing a theoretical opportunity for the secondary gene rearrangements.

In order to better understand whether *IGLV* gene families could be involved in secondary gene rearrangements in the B cell subsets studied, the relative utilization of *IGLJ* gene segments by different *IGLV* gene families was determined. Although there were no statistical differences in *IGLJ* usage, there were different trends of involvement of *IGLJ* gene segments with different *IGLV* gene families. For example, there tended to be differences in involvement of *IGLJ*1 in naïve as compared to IgA expressing B cells as there was lower frequency in IgA expressing B cells as compared to the naïve B cells that could be consistent with light chain revision in IgA (Figure 3-21).

.

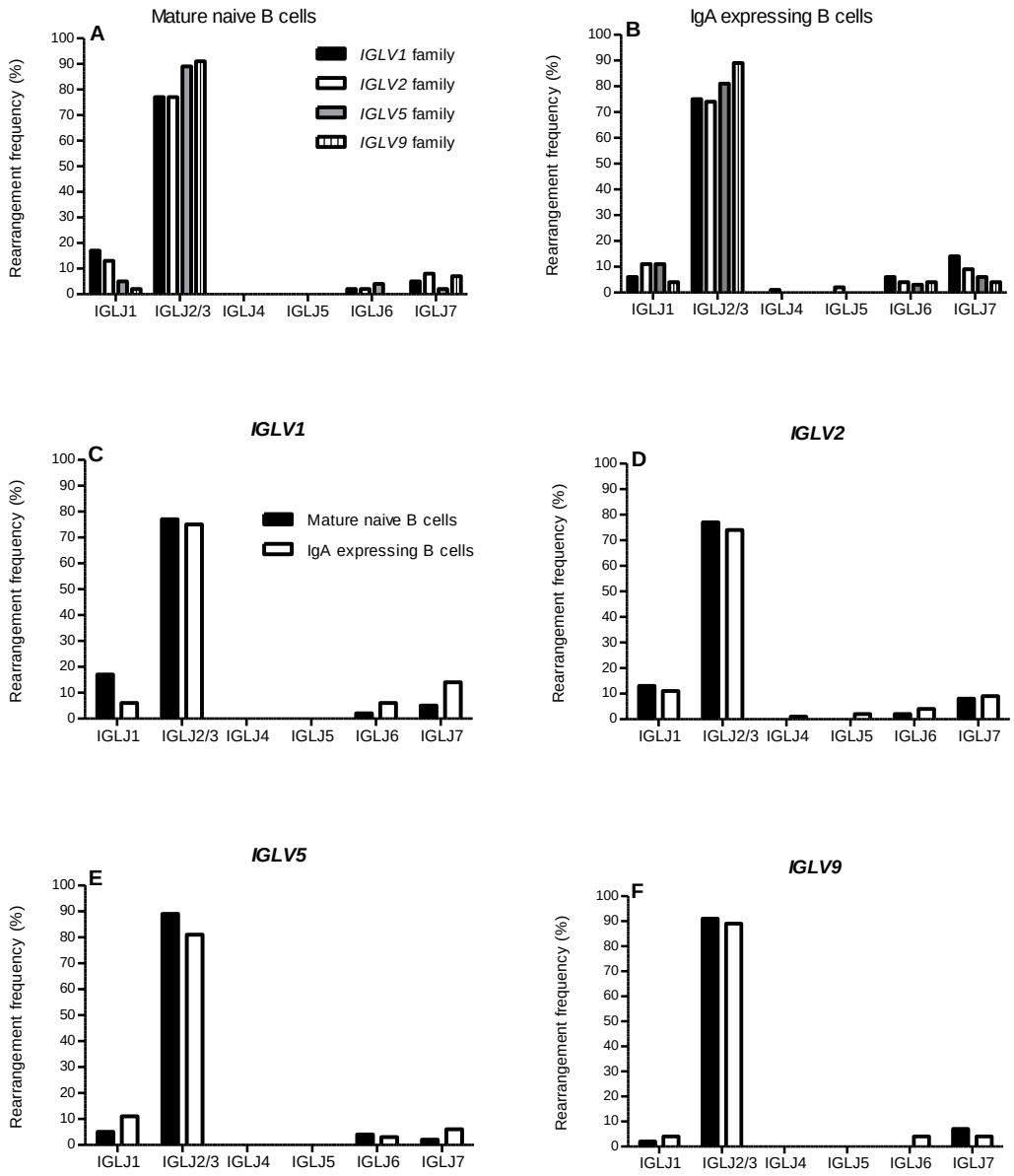
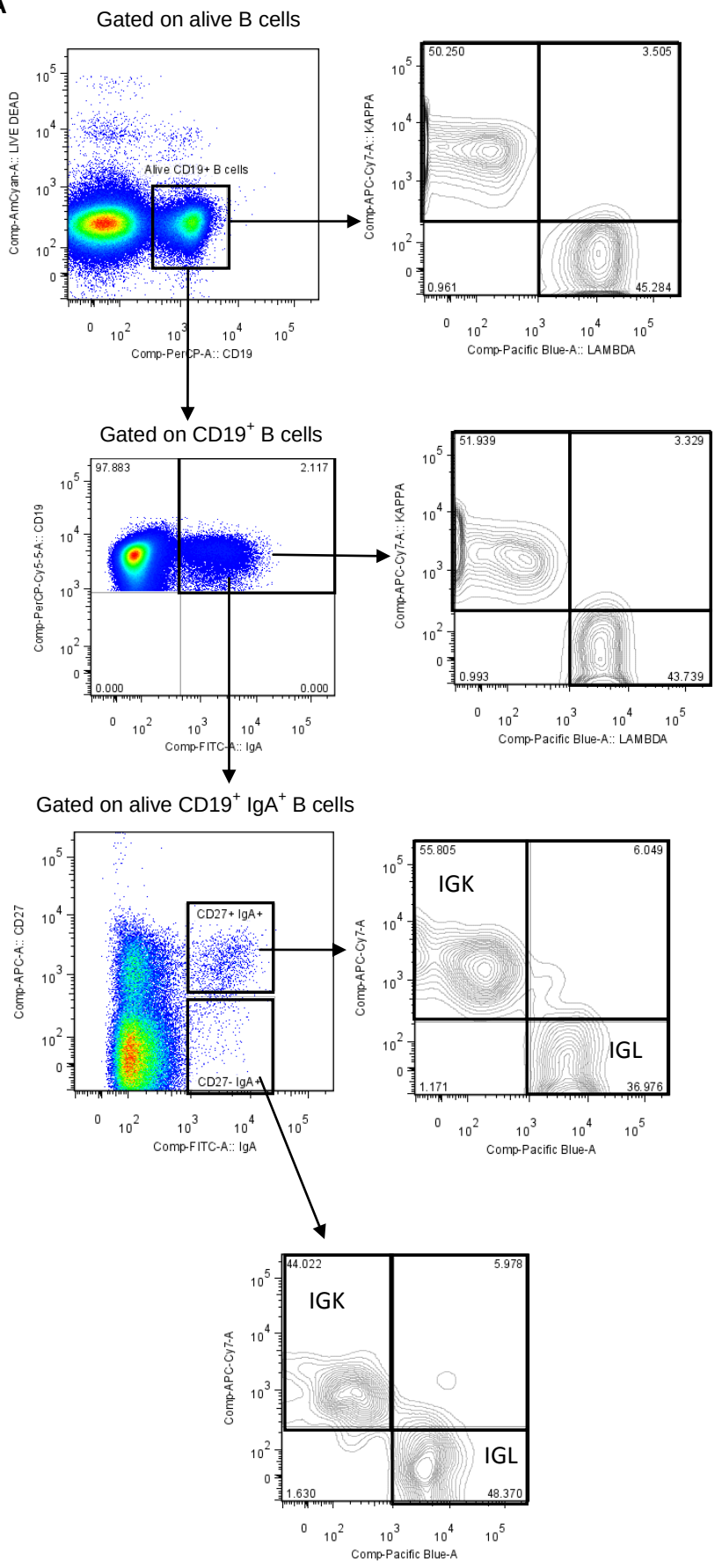


Figure 3-21: Differences in the rearrangement frequencies of various *IGLJ* segments with different *IGLV* gene families. The relative gene rearrangement frequencies of various *IGLV* families with different *IGLJ* gene segments naïve and IgA expressing B cells has been shown in **A** and **B**. **C** to **F** are showing relative gene rearrangement frequencies of *IGLV1*, *IGLV2*, *IGLV5* and *IGLV*9 families in naïve and IgA expressing B cells. There was reduction in *IGLJ*1 usage in IgA expressing B cell (not statistically significant). There was similar usage of *IGLJ* gene segments with *IGLV*2 in naïve and IgA expressing B cells. There was negligible use of *IGLJ*1 in the gene rearrangements involving *IGLV*9 in both naïve and IgA.

3.2.6 Is there any evidence of a bias in light chain expression in IgA by flow cytometry?

It has previously been suggested that there are more IGL expressing B cells as compared to IGK expressing B cells in $CD27^-$ IgA B cells than $CD27^+$ IgA B cells (Berkowska, Driessen et al. 2011). Therefore, it was decided to compare IGK to IGL ratios in different subsets of IgA to find out if the proportion of IGK and IGL varies. For this purpose, blood was collected from 10 buffy coats and stained with CD19-PerCPcy5.5, CD27- APC, IgA-FITC, IGK-APC-cy7 and IGL-Pacific blue (as described in the section 2.3.5). The data was acquired using flowcytometer and analyzed with Flowjo software. Consistent with previous study, it was found that in $CD27^-$IgA B cells there was a greater proportion of IGL of total B cells as compared to $CD27^+$IgA B cells (Figure 3-22).

A


Gated on alive B cells
Comp-AmCyan-A:: LIVE DEAD
Alive CD19+ B cells
Comp-PerCP-A:: CD19
Comp-APC-Cy7-A:: KAPPA
50.250
3.505
0.961
45.284
Comp-Pacific Blue-A:: LAMBDA
Gated on CD19+ B cells
Comp-PerCP-Cy5-5-A:: CD19
97.883
2.117
0.000
0.000
Comp-FITC-A:: IgA
Comp-APC-Cy7-A:: KAPPA
51.939
3.329
0.993
43.739
Comp-Pacific Blue-A:: LAMBDA
Gated on alive CD19+ IgA+ B cells
Comp-APC-A:: CD27
CD27+ IgA+
CD27- IgA+
Comp-FITC-A:: IgA
Comp-APC-Cy7-A
55.805
6.049
IGK
IGL
1.171
36.976
Comp-Pacific Blue-A
44.022
5.978
IGK
IGL
1.630
48.370
Comp-Pacific Blue-A

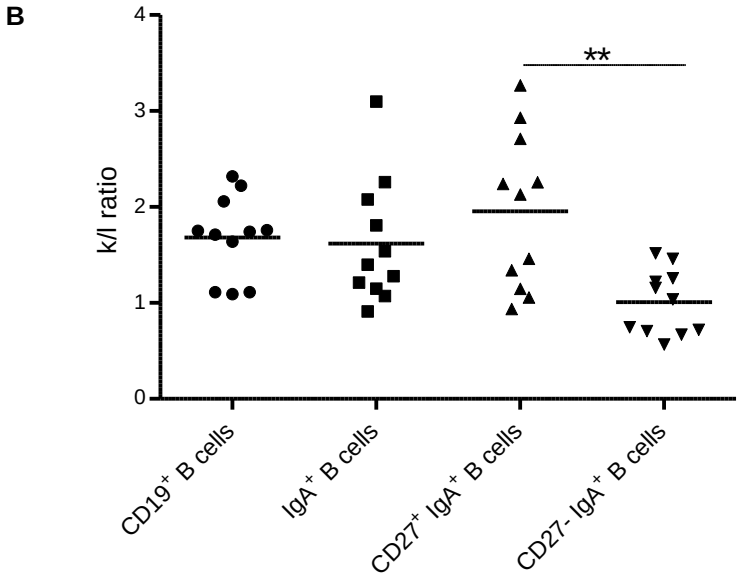


Figure 3-22: Differences in kappa to lambda ratios between CD27^{+}IgA and CD27^{-}IgA B cells. The gating strategy to identify CD27^{+}IgA and CD27^{-}IgA B cell subsets has been shown (**A**). The scatter plot showing mean of IGK/IGL in CD27^{+}IgA and CD27^{-}IgA B cell subsets. Non-parametric (Mann-Whitney) test was used to compare the means (**B**). The IGK:IGL is significantly decreased in the CD27^{-}IgA B cells.

3.3 Discussion

3.3.1 More out-of-frame than in-frame gene rearrangements in mature naïve B cells

In the data studied here almost 55% of total *IGL* gene rearrangements amplified from DNA extracted from naïve B cells were out-of-frame. The abundance of out-of-frame gene rearrangements in the repertoire of naïve B cells could suggest that receptor editing has occurred so that a cell can carry more than two rearrangements. The primary V(D)J gene rearrangement could generate a CDR-3 junction having either in-frame or out-of-frame functionality. If the rearrangement is autoreactive in the first place then it can also be replaced by editing and new rearrangement could be out-of-frame thus further adding to total out-of-frame gene rearrangements. When the junction is out-of-frame then it leads to the initiation of the receptor editing either on the same allele or second allele in an attempt to generate a functional BCR (Luning Prak, Monestier et al. 2011). As a result of the secondary gene rearrangement, specificities of BCRs can be modified which could be beneficial because it causes diversification of repertoire and elimination of autoreactive specificities (Nemazee 2006). However, non productive receptor editing defaults to clonal deletion (Halverson, Torres et al. 2004). Overall this editing process generates episomal fragments of deleted out-of-frame gene rearrangements that could be amplified by PCR resulting in the potential to accumulate more out-of-frame gene rearrangements in next generation sequencing data than may have been predicted. Relative reduction in the proportion of out-of-frame rearrangements in IgA could be generated by cell division that would replicate the chromosomes but not the episomes generated by secondary rearrangement events of editing process.

It is also possible that some out-of-frame rearrangements might be generated artificially by the 454 sequencing method that is known to insert nucleotides at homopolymer repeats. This error may enhance the number of out-of-frame rearrangements observed since by chance; there is 2 times the possibility of rearrangements being out-of-frame than in-frame. However, there were consistent observations in the IgA subsets studied that were different to the mature naïve B cells that would have been affected equivalently by this, suggesting that this did not make a major impact on the data.

3.3.2 Gene families *IGLV5* and *IGLV9* are skewed towards out-of-frame gene rearrangements in mature naïve and IgA expressing B cells

Unlike *IGLV*1 and 2 families, gene rearrangements involving *IGLV*5 and 9 gene families were mostly out-of-frame. This is consistent with previous study, however, abundance in the out-of-frame gene rearrangements was observed not only in IgA expressing B cells but also in the mature naïve B cells (Farner, Dorner et al. 1999, Su, Boursier et al. 2004). Gene families *IGLV*5 and *IGLV*9 are in the cluster B and C which increases the likelihood of involvement of these families in the secondary gene rearrangements. In this data it was found that *IGLJ*1 was not involved in the gene rearrangements involving *IGLV*9 family. It may suggest the involvement of this family in receptor editing and revision. It has been reported there is a stop codon at the end of *IGLV*9 region. Although rare, nonsense codons at the end of *IGLV* region limit functional rearrangement for these *IGLV* genes (Kawasaki, Minoshima et al. 1997, Schoettler, Ni et al. 2012). Moreover, *IGLV*9-49 is also known to exhibit intrinsic autoreactivity which may subject these rearrangements to negative selection before expression (Schoettler, Ni et al. 2012). However, which factors are involved for the abundance of out-of-frame gene rearrangements involving *IGLV*5 family are not known. All these gene rearrangements were studied at the level of DNA. Analysis of cDNA gene rearrangements in various B cell subsets revealed that *IGLV*5 and 9 were not used to constitute BCR at any stage of the B cell development.

3.3.3 Is there any evidence for receptor revision of IgA in this analysis?

Previous studies of cell lines derived from mucosal B lineage cells and features of light chain excision circles have suggested that the repertoire of mucosal IgA plasma cells can be modified by receptor revision (Su, Gordon et al. 2008). Re expression of *RAG* genes has been reported *in vitro* in mature peripheral B cells by stimulation with SAC, IL-2 and also with anti-CD40 antibody and IL-4 (Nagafuchi, Yoshikawa et al. 2004). Re-expression of RAGs in human mature B cells has also been reported and the purpose was suggested to increase diversification of the immune repertoire (Han, Dillon et al. 1997, Hertz, Kouskoff et al. 1998, Hikida and Ohmori 1998). Reactivation of RAG genes has also been found in mature peripheral B cells isolated from the tonsils of humans and mice (Meffre, Papavasiliou et al. 1998). Mature B cells can re-induce RAG genes after somatic mutation in order to

revise the BCR by secondary recombination for diversification and to rescue B cells from apoptosis by re arranging light chains (Han, Zheng et al. 1996, Nemazee and Weigert 2000). The transcripts of *rag1* and *rag2* genes and proteins have been documented in human GC B cells and in extra follicular areas (Giachino, Padovan et al. 1998, Girschick, Grammer et al. 2001, Meru, Jung et al. 2002). In physiological conditions receptor revision has been proposed as a mechanism for receptor diversification to increase the diversity of BCRs (Meffre, Papavasiliou et al. 1998). Expression of TdT has also been found in mature B cells in the periphery to add random nucleotides at the junction and thus diversification of repertoire (Komori, Okada et al. 1993).

No evidence of receptor revision was observed in this study however since no significant biases in *IGLJ* segment usage were observed in any B cell subset analysed or any *IGLV* segments used by them. It is known that IgA producing cells in blood are not all mucosal and mucosal IgA producing cells were not studied in this thesis. It would be interesting to know if the mucosal B cell subset in blood that expresses higher levels of α4β7 integrin would have generated a different profile.

3.3.4 CD27$^-$IgA B cells have lower IGK/IGL ratio than CD27$^+$IgA B cells

It is possible that CD27$^+$ IgA and CD27$^-$ IgA represent distinct subsets of IgA memory B cells. It has been suggested that CD27$^-$ IgA memory B cells could be generated as a result of TID responses in the gut because of the presence of these cells in CD40L deficient patients. TID IgA responses have been observed both in human and mouse, both systemically in the splenic marginal zone and locally in the gastrointestinal system (Fagarasan, Kinoshita et al. 2001, Bergqvist, Gardby et al. 2006). Potential mediators of CD40-independent IgA CSR are BAFF and APRIL (Litinskiy, Nardelli et al. 2002). High frequency of IGL expressing B cells in CD27$^-$IgA is consistent with the previous study that suggested a broad IGL repertoire may be beneficial for responses in the human gastrointestinal tract (Berkowska, Driessen et al. 2011). It gives a clue that IGK and IGL expression changes during developmental stages which conceptually could be observed by the initiation of *IGL* gene rearrangements in IGK expressing B cells or on the previously unrearranged *IGL* locus. No differences in the profile of *IGL* repertoire was observed in this

thesis between the $CD27^+$ and $CD27^-$ subsets of IgA according to any of the parameters studied. Despite this and although it has been shown that the same clone of B cells can be present in both $CD27^+$ and $CD27^-$ IgA expressing B cell subsets, the consistent differences in IGK:IGL ratio between studies is consistent with the idea that the these subsets of IgA producing cells may have a different derivation.

3.4 Conclusions

1. No bias was observed towards in-frame gene rearrangements involving *IGLV*1 and *IGLV*2 gene families in transitional B cells as observed previously in IgA secreting cells.
2. There were more in-frame gene rearrangements overall in IgA e*xpre*ssing B cells as compared to mature naïve B cells. This trend was consistent in all subsets of IgA expressing B cells: IgA plasmablasts, $CD27^+$ IgA and $CD27^-$IgA.
3. There were no differences in the profile of out-of-frame gene rearrangements between naïve and IgA expressing B cells.
4. The gene segment *IGLV*2-14 is preferentially used more in IgA expressing B cells as compared to mature naïve B cells.
5. There is no evidence of the secondary rearrangement as revealed by the comparison of *IGLJ* gene segments. However, there was a trend of less usage of *IGLJ*1 by IgA expressing B cells as compared to naïve B cells.
6. There is a bias towards more IGL expressing cells in $CD27^-$IgA than in $CD27^+$IgA.

Chapter 4

High throughput sequencing of the human immunoglobulin kappa light chain gene rearrangements of mature naïve B cells

4.1 Introduction

B cell development is characterized by a well defined order of rearrangements at *IGH*, *IGK* and *IGL* chain loci. Somatic recombination of *IGHD* and *IGHJ* genes is followed by recombination of *IGHV* and rearranged *IGHD-J* segments. Gene rearrangement involves removal of nucleotides by exonucleases and addition of nucleotides by TdT. Hence there is only a 1/3 chance of having a correct genetic reading frame and 2/3 attempts would yield out of frame gene rearrangements (Hardy and Hayakawa 2001, Santos, Arumemi et al. 2011). If the rearrangement is productive (in frame with no stop codons) this newly synthesized μHC is expressed on the cell surface with SLC (VpreB and λ5) and transmembrane proteins Igα and Igβ. Expression of this pre-BCR on the cell surface is considered a critical signal for *IGH* exclusion of the second allele. If rearrangement is not successful, the process is repeated on the second allele. Upon pre-BCR signalling cells undergo proliferation and there is down regulation of RAG and SLC. Signalling also induces transcription factors such as NF-κB, Spi-B and IRF-4 as well as re-expression of RAG that results in the activation of κ enhancers that initiates rearrangement at the *IGK* locus (van Zelm, van der Burg et al. 2005).

In humans there are about 40 (out of 76) functional *IGKV* gene segments and 5 *IGKJ* gene segments (Figure 1-4). Rearrangement proceeds by alignment of one *IGKV* allele to *IGKJ* on one allele. The productive gene rearrangements may be translated into protein that could be expressed as part of a functional antigen receptor. B cells expressing surface immunoglobulin can be influenced by a variety of selective forces resulting in their preferential expansion or deletion from the repertoire. However, if the genetic rearrangement is non productive due to being out of frame or having stop codons it would not be translated into a protein; their presence in the cell does not influence the fate of the cell. Therefore these gene rearrangements reflect intrinsic biases in factors such as transcriptional activity or RSS sequence that influence the rearrangement process (Foster, Brezinschek et al. 1997). If gene rearrangement is not successful then the *IGK* locus is inactivated by rearrangement of the KDE and gene rearrangement process of *IGL* locus starts (van der Burg, Tumkaya et al. 2001, Vela, Ait-Azzouzene et al. 2008). KDE is non-coding DNA located 25 kb downstream of *IGKC*. The RSS heptamer located 5’ of the KDE is separated from the nonamer by a 23bp sequence. The RSS configuration resembles that found 5’ of

each *IGKJ* segment. The target of the KDE is an RSS composed of a heptamer separated from a nonamer by 30 bp spacer. There are two ways of KDE rearrangement. It can undergo RSS mediated site specific recombination with intronic RSS located between distal *IGKJ* and *IGKC*. On the other hand KDE can also recombine with the RSS of an upstream *IGKV* segment that has not been rearranged. As a result of these mechanisms there is a deletion of *IGKC*, iEk and 3'Ek enhancers (Siminovitch, Bakhshi et al. 1985, Nayar, Paklet et al. 2010). Consequently there is inactivation of *IGK* locus by deletion. It has also been found that iRSS can also inactivate the *IGK* locus independently with an intermediate step of rearrangement to *IGKJ* downstream to the non productive gene rearrangement (Feddersen, Martin et al. 1990, Feddersen, Martin et al. 1990, Langerak, Nadel et al. 2004).

Gene rearrangement proceeds in an ordered fashion and is initiated at *IGL* locus if both alleles of *IGK* locus fail to produce a functional protein. Interestingly it has been found that non productive gene rearrangement in terms of DNA sequence at the *IGK* locus is not the only factor to initiate gene rearrangement at *IGL* locus. Approximately 30% of mature naïve B cells expressing IGL on the surface had undergone productive gene rearrangement at the *IGK* locus already according to the sequence of the rearranged allele. However, these rearrangements were not selected to constitute functional BCR rendering them non productive in functional terms (Brauninger, Goossens et al. 2001). These productive rearrangements of *IGK* in IGL expressing B cells are referred to as aberrant productive gene rearrangements in this thesis. They are of interest because despite being apparently functional according to sequence they have been selected against and are not used to generate BCR.

One of the characteristic features of the *IGK* locus is the orientation of *IGKV* gene segments. Most of the *IGKV* gene segments like *IGKJ* segments are in the reverse orientation. However, the rest of the *IGKV* segments are in the forward orientation as shown in Figure 1-4 (Foster, Brezinschek et al. 1997, Schoettler, Ni et al. 2012). Of note, in this thesis *IGKV*4-1 and *IGKV*5-2 are in the forward orientation. Complete classification of *IGKV* genes according to the orientation on chromosome has been given in Figure 1-4. The different orientation of *IGKV* ultimately determines the consequences of rearrangement. If *IGKV* gene segments

are in the same orientation as *IGKJ* gene segments then alignment of the two RSS regions loops out the intervening DNA. The intervening DNA is excised from the chromosome and cannot be used for further rearrangement (Figure 4-1). However, when *IGKV* and *IGKJ* gene segments are in the opposite orientation, alignment of the RSS regions form the intervening DNA into a coiled configuration. As a result after recombination the coiled region is retained in the chromosome in an inverted orientation (Figure 4-2).

Various biases have been reported in the usage of *IGKV* and *IGKJ* segments in the B cell repertoire (Jackson, Kidd et al. 2013). However, comparison of productive rearrangements with expressed rearrangements has not been reported so far. It is noteworthy that not all productive rearrangements are used to encode BCR.

The experiment in this chapter is designed to allow analysis of selection that may occur after gene rearrangement and before maturation to mature naïve B cells. In this study 5 groups of sequences were analyzed.

1. Productive DNA gene rearrangements
2. Non productive DNA gene rearrangements
3. Expressed cDNA gene rearrangements

From IGK expressing mature naïve B cells

4. Aberrant productive DNA gene rearrangements
5. Non productive DNA gene rearrangements

From IGL expressing mature naïve B cells

The schematic representation of the experiment and functional relevance of different rearrangements from different subsets is given in Figure 4-3. In addition to *IGK* genes usage CDR-3 characteristics were also compared. The CDR-3 region is a major contributor to antigenic specificity. The physicochemical properties of the CDR-3 region such as CDR-3 length, amino acids distribution, GRAVY index, aliphatic index, isoelectric points were also compared among non productive, aberrant productive, productive and expressed sequences.

4.1.1 Aims of this chapter

The aims of investigation in this chapter were to ask:

1. Does the expressed *IGKV* gene repertoire mimic the productive DNA gene rearrangements in IGK expressing B cells? This will be addressed by comparing productively rearranged DNA and cDNA from IGK producing B cells.
2. Does the expressed *IGKJ* gene repertoire reflect the productive DNA gene rearrangements in IGK expressing B cells? This will be addressed by comparing productively rearranged DNA and cDNA from IGK producing B cells.
3. What are the inherent biases during DNA gene rearrangement process at *IGK* locus? This will be visualized by studying non productively unselected gene rearrangements.
4. Is there any selection for or against of *IGKV* gene segments during rearrangement and expression? This will be studied by comparing productive gene rearrangements from IGK and IGL expressing mature naïve B cells.
5. Is there any difference in the characteristic of CDR-3 region of aberrant productive, productive and expressed repertoire? For this purpose following characteristics of CDR-3 region were analyzed:
    i. CDR-3 length
    ii. Relative distribution of amino acids
    iii. GRAVY indices
    iv. Isoelectric points
    v. Aliphatic indices

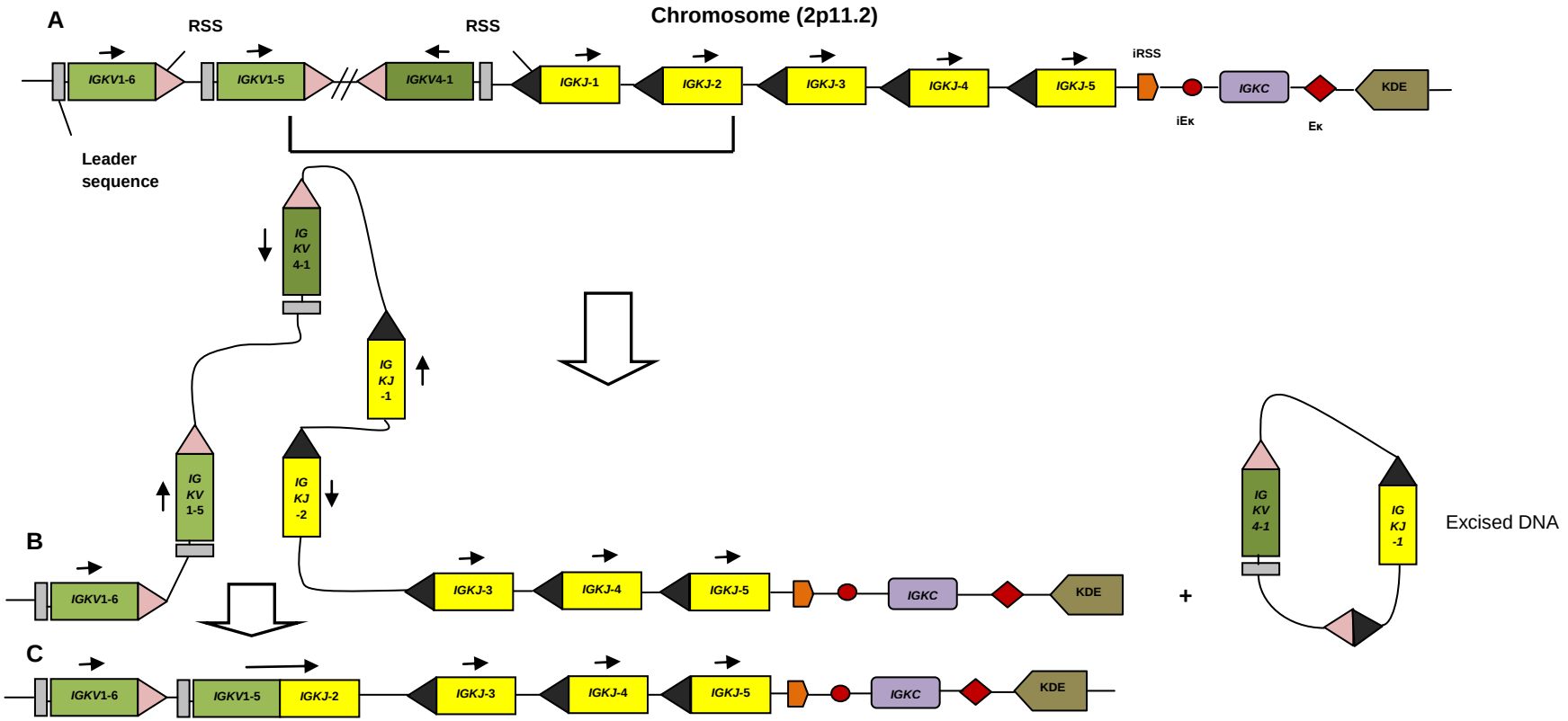


Figure 4-1: Schematic representation of *IGKV* to *IGKJ* gene rearrangement having same orientation. (A) Various gene loci are shown on long arm of chromosome 2. *IGKV* gene segments (green), *IGKJ* gene segments (yellow), intronic RSS (brown), enhancers (red), constant region (purple) and KDE (olive green). Arrows indicate the direction of transcription. **(B)** When *IGKV* and *IGKJ* gene segments are in the same orientation there is juxtaposition and subsequently excision of intervening DNA (excised DNA) and RSS of the rearranged genes. In this example rearrangement of *IGKV*1-5 with *IGKJ*-2 has been shown. **(C)** After excision of intervening DNA, *IGKV*1-5 and *IGKJ*-2 has formed a coding joint which can be transcribed.

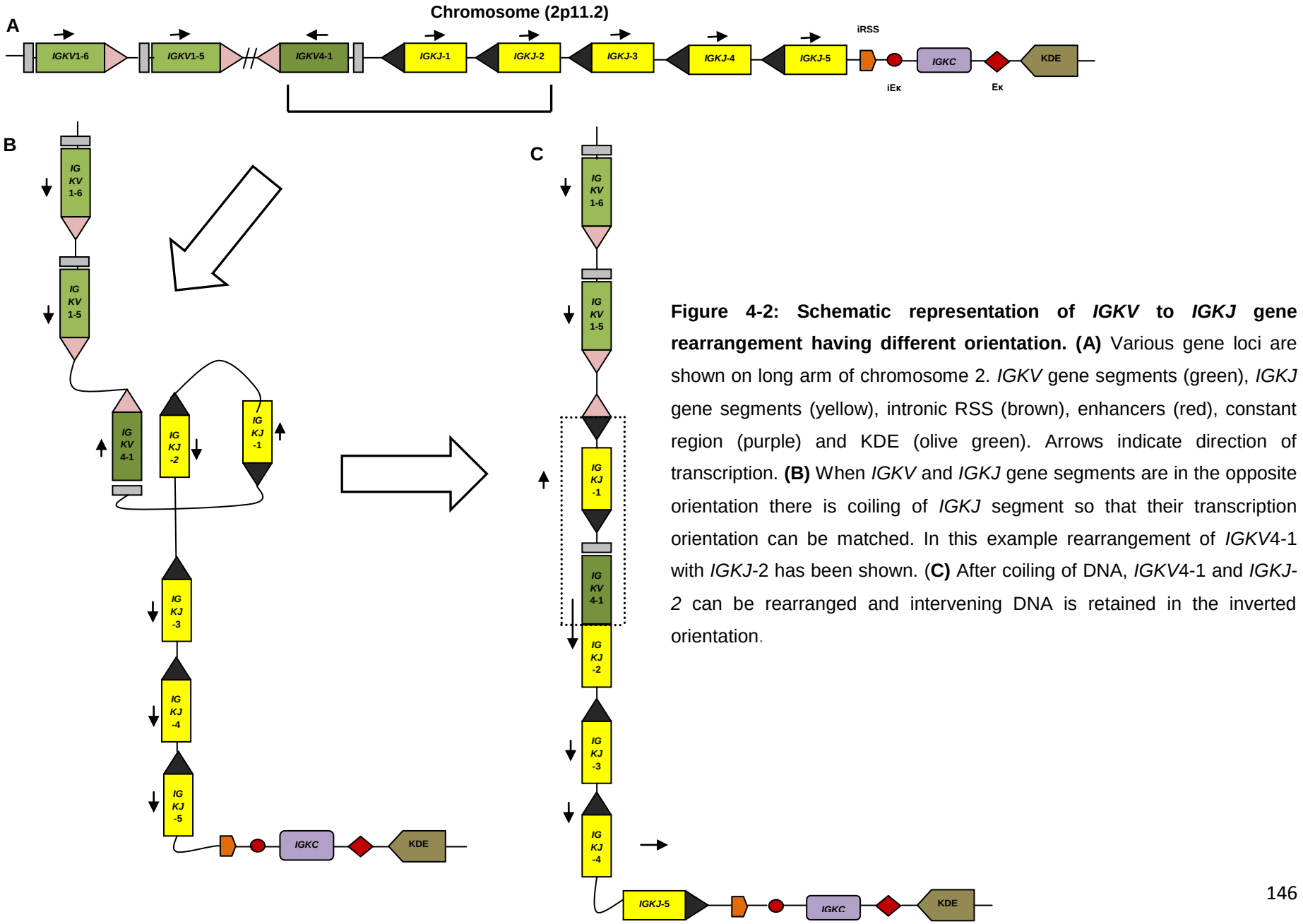


Figure 4-2: Schematic representation of *IGKV* to *IGKJ* gene rearrangement having different orientation. (A) Various gene loci are shown on long arm of chromosome 2. *IGKV* gene segments (green), *IGKJ* gene segments (yellow), intronic RSS (brown), enhancers (red), constant region (purple) and KDE (olive green). Arrows indicate direction of transcription. **(B)** When *IGKV* and *IGKJ* gene segments are in the opposite orientation there is coiling of *IGKJ* segment so that their transcription orientation can be matched. In this example rearrangement of *IGKV*4-1 with *IGKJ*-2 has been shown. (**C)** After coiling of DNA, *IGKV*4-1 and *IGKJ-2* can be rearranged and intervening DNA is retained in the inverted orientation.

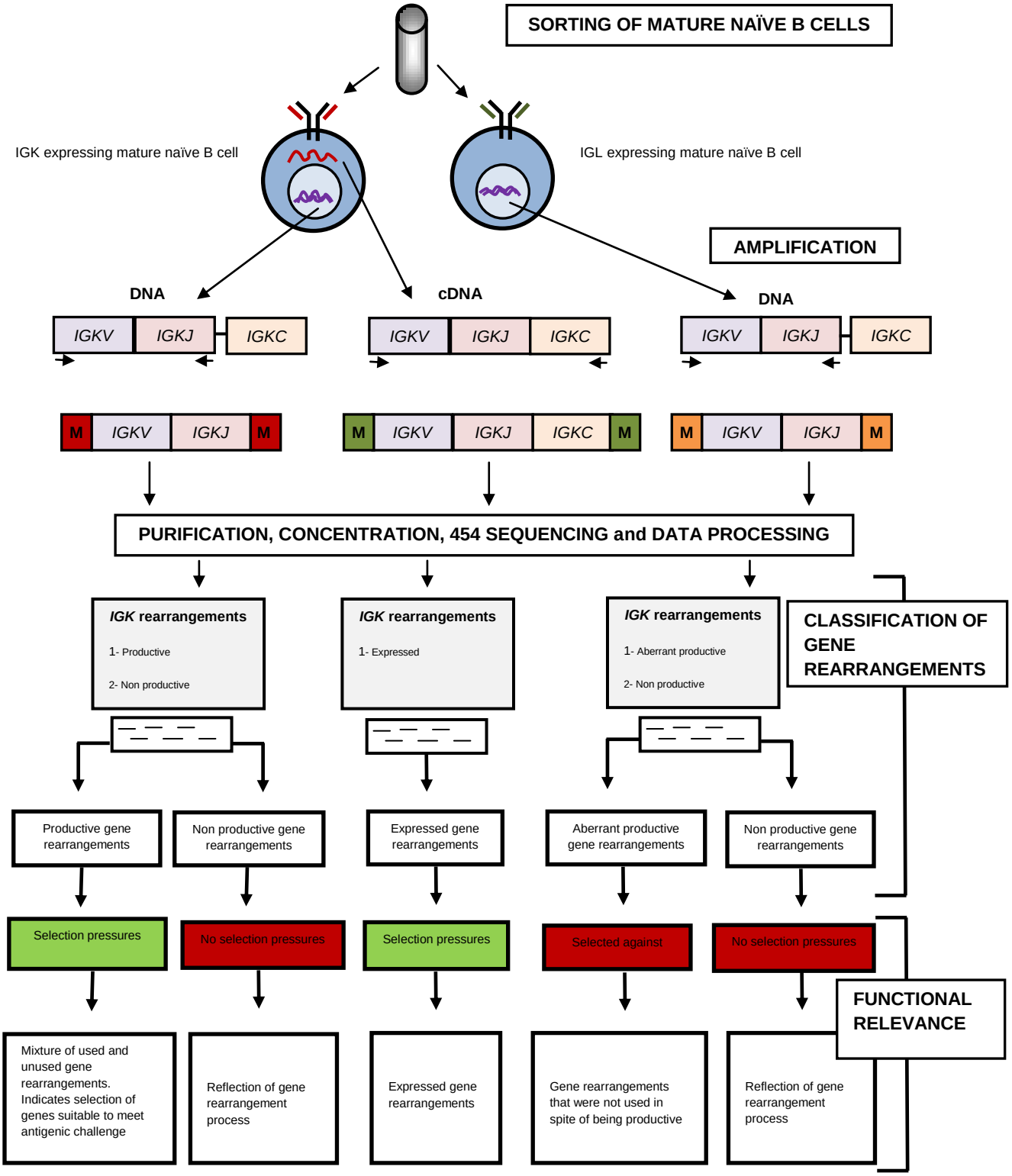


Figure 4-3: Flow chart showing various stages of experiment and data interpretation. The PBMCs from three healthy individuals were sorted into mature naïve B cells according to surface light chain expression. Genomic DNA and cDNA was extracted from IGK expressing and genomic DNA from IGL expressing mature naïve B cells and amplified (arrows indicate binding sites of the primers). Sequences were retrieved and passed through quality control. Functional relevance of different sets of gene rearrangements is shown.

4.2 Results

The PBMCs were isolated from the blood of three healthy individuals. Cells were stained with flurochrome tagged monoclonal antibodies against CD19-PerCPcy5.5, CD27-APC, IgD-PE-cy7, kappa-APC-cy7, lambda-Pacific blue and live/dead-Amcyan. The gating strategy for the identification of cells is given in Figure 4-4.

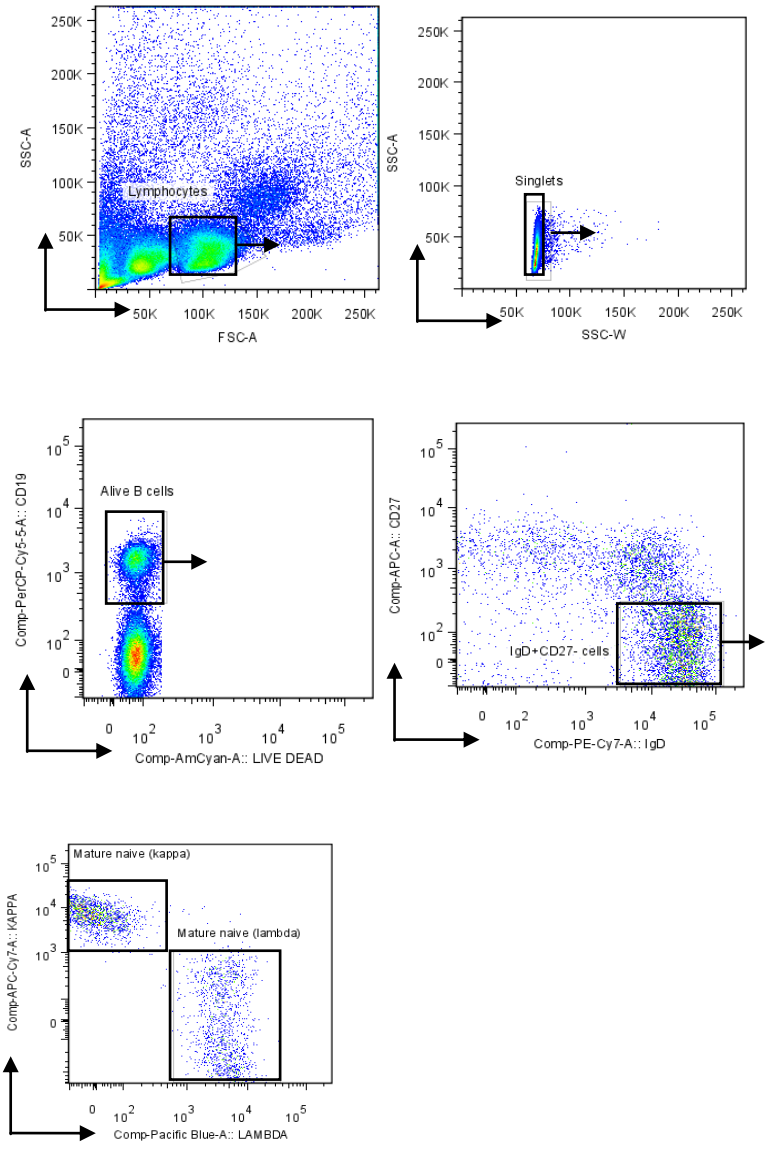


Figure 4-4: Example of sorting B cell subsets into mature naïve B cells expressing either IGK or IGL light chain. PBMCs were stained with monoclonal antibodies and sorted using FACS. Lymphocytes were gated first and doublets were excluded. Cells were first gated on alive $CD19^+$ B cells and then on IgD^+CD27^-. Mature naïve B cells ($CD19^+CD27^-IgD^+$) were divided according to surface expression of light chains (IGK or IGL).

Mature naïve B cells expressing either IGK ($CD19^+CD27^-IgD^+IGK^+$) or IGL ($CD19^+CD27^-IgD^+IGL^+$) were sorted in to separate tubes containing RPMI medium supplemented with FCS and antibiotics (penicillin and streptomycin). Mature naïve B cells expressing IGK were divided into two fractions for the extraction of DNA (genomic repertoire) and cDNA (expressed repertoire). In order to analyze the *IGK* rearrangements carried by mature naïve B cell expressing IGL only genomic DNA was extracted. The number of cells used to acquire the sequences is given in Table 4-1.

Table 4-1: The respective number of mature naïve B expressing either IGK or IGL derived from each donor

Samples	Mature naïve B cells expressing IGK	Mature naïve B cells expressing IGL
HD-1	65,955	37,479
HD-2	274,790	204,037
HD-3	362,997	296,590

For the amplification of rearranged (genomic) and expressed (cDNA) repertoire, primers were used from *IGKV* to *IGKJ* and *IGKV* to *IGKC* respectively. In order to distinguish amplified products from different individuals for the second round of PCR, primers were tagged with different barcodes consisting of 10-nucleotides sequence of MID. Different tags that were used to segregate amplified products are given in Table 4-2.

Table 4-2: The primer tags used for differentiating the amplified products from different individuals

Sample	MID tags used to distinguish products
HD-1	IGK expressing mature naïve B cells (DNA) MID-1
	IGK expressing mature naïve B cells (cDNA) MID-2
	IGL expressing mature naïve B cells (DNA) MID-3
HD-2	IGK expressing mature naïve B cells (DNA) MID-5
	IGK expressing mature naïve B cells (cDNA) MID-6
	IGL expressing mature naïve B cells (DNA) MID-7
HD-3	IGK expressing mature naïve B cells (DNA) MID-9
	IGK expressing mature naïve B cells (cDNA) MID-10
	IGL expressing mature naïve B cells (DNA) MID-11

Excess primers were removed after running on agarose gel. Amplified products of required sizes (genomic: 350bp, cDNA: 450bp) were excised from the agarose gel and purified and quantified using Qubit fluorometer. Samples were pooled and concentrated. Sample was sequenced using GS FLX Genome Titanium Sequencer 454. Resulting sequences were passed through quality control criteria (section 2.3.6.5).

From three donors collectively 1,027 unique gene rearrangements were acquired from genomic DNA (*IGKV-JK*) and 4,360 gene rearrangements from cDNA (*IGKV-CK)* of IGK expressing mature naïve B cells. From mature naïve B cells expressing IGL on the cell surface 2,459 genomic DNA (*IGKV-JK)* gene rearrangements were obtained. Based on CDR-3, *IGKV*, I*GKJ* and *IGKC* genes usage only one gene rearrangement was selected in the analysis (Table 4-3).

Gene rearrangements from genomic DNA of mature naïve B cells expressing IGK on the cell surface were further divided into productive and non productive sequences according to the functionality of the junction. If the junction between *IGKV* to *IGKJ* was in the correct genetic

reading frame with no stop codons, rearrangements were classified as **productive sequences**. If they were out of frame or had stop codons they were classified as **non productive sequences**. In order to analyze the expressed IGK repertoire of mature naïve B cells, cDNA was also isolated. Non productive *IGKV-CK* gene rearrangements were excluded from the data analysis.

Rearrangements of *IGK* from mature naïve B cells expressing IGL on cell surface were also divided according to the functionality of junction. As these gene rearrangements were from cells that expressed IGL not IGK on the cell surface they were not used to constitute functional BCR in these cells. Therefore, such productive DNA gene rearrangements in IGL expressing B cells were classified as **aberrant productive sequences**. Like IGK expressing B cells, if the genetic reading frame was incorrect or had stop codons such sequences were classified as **non productive sequences**. The percentages of aberrant productive, productive and non-productive gene rearrangements from each cell type are shown in Figure 4-5. The relative number of sequences derived from each donor for each subset is given in the Table 4-3.

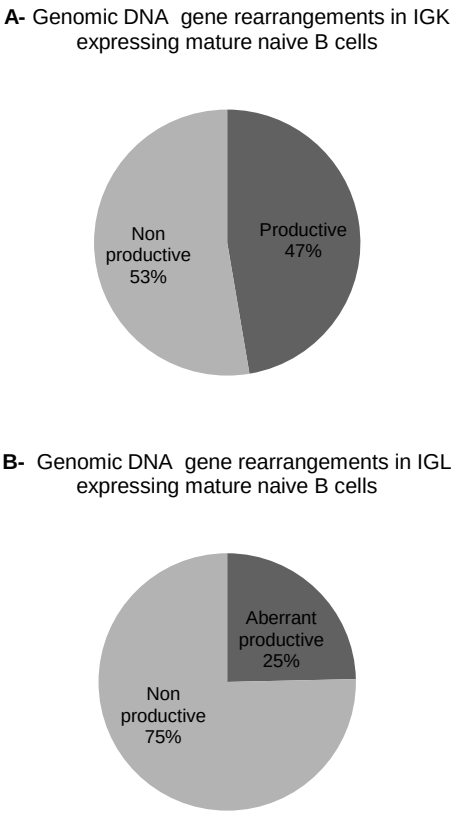


Figure 4-5: Classification and respective percentages of genomic DNA gene rearragments at *IGK* locus obtained from mature naïve B cells expressing expressing either IGK **(A)** or IGL **(B)** on the cell surface**.**

Table 4-3: The respective number of unique gene rearrangements derived from mature naïve B cells (sorted according to surface light chain expression) from three healthy donors (HD)

	***VJC* gene rearrangements**	HD-1	HD-2	HD-3	**Total**
1	Productive DNA gene rearrangements from IGK expressing mature naïve B cells	329	81	76	486
2	Non productive DNA gene rearrangements from IGK expressing mature naïve B cells	364	124	53	541
3	Expressed gene rearrangements from IGK expressing mature naïve B cells (cDNA)	1071	2787	502	4360
4	Aberrant productive DNA gene rearrangements from IGL expressing mature naïve B cells (DNA)	469	28	109	606
5	Non productive DNA gene rearrangements from IGL expressing mature naïve B cells	1455	97	301	1853

4.2.1 Does the expressed *IGKV* gene repertoire resemble the productive DNA gene rearrangements?

In order to create the light chain of the BCR, gene rearrangement starts first at the *IGK* locus at the pre B cell stage during B cell development in the bone marrow. Productive gene rearrangements are translated into functional protein. Selection pressures are imposed on the productive gene rearrangements in the bone marrow and some of the productive rearrangements are not ultimately used to constitute BCR. Furthermore, receptor editing by secondary gene rearrangements may occur if gene rearrangement at first allele rearrangement is selected against. As a consequence, productive gene rearrangements may accumulate in a cell and may not all encode protein. To test this, the relative usage of *IGKV* families and *IGKV* gene segments was compared between productive DNA gene rearrangements with the expressed DNA gene rearrangements.

Differences were observed between the profile of expressed genes and the profile of productive gene rearrangements in DNA. It was found that overall rearrangements using the *IGKV*3 family were relatively more in cDNA as compared to DNA. On the other hand *IGKV*4 family was significantly lower in frequency in the expressed repertoire (Figure 4-6). This was true even when individual donors were analyzed separately (data not shown). This suggests that productive gene rearrangements of *IGKV*4 family accumulate in a cell but are not expressed and may be selected against.

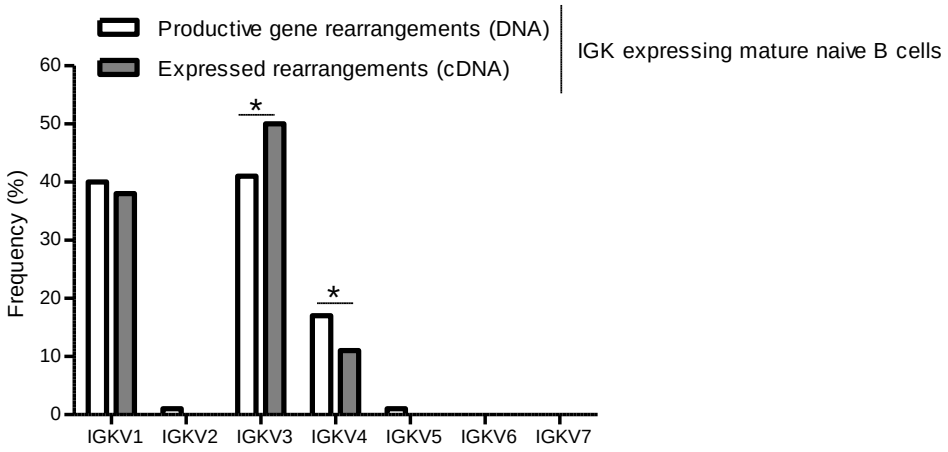


Figure 4-6: Distribution of *IGKV* families in the productive and expressed gene rearrangements. Comparison of relative rearrangement frequencies of *IGKV* gene families between productive and expressed gene rearrangements of IGK expressing mature naïve B cells. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). *IGKV*3 was significantly increased while *IGKV*4 was significantly reduced in the expressed repertoire.

The relative frequency of *IGKV* gene segments was then compared between productive rearrangements in DNA with the expressed gene rearrangements from cDNA. The gene segments *IGKV*3-15 and *IGKV*3-20 were found significantly more in the expressed repertoire. In contrast to this, rearrangements using *IGKV*4-1 rearrangements were significantly more abundant in DNA as compared to cDNA (Figure 4-7). In order to validate the pooled data the sequences from each individual were analyzed separately. It was found that in all individuals there was more utilization of at least one *IGKV*3 family member although not the same in each case. *IGKV*4-1 was consistently less abundant in the cDNA relative to the productive rearrangements in DNA (Figure 4-8).

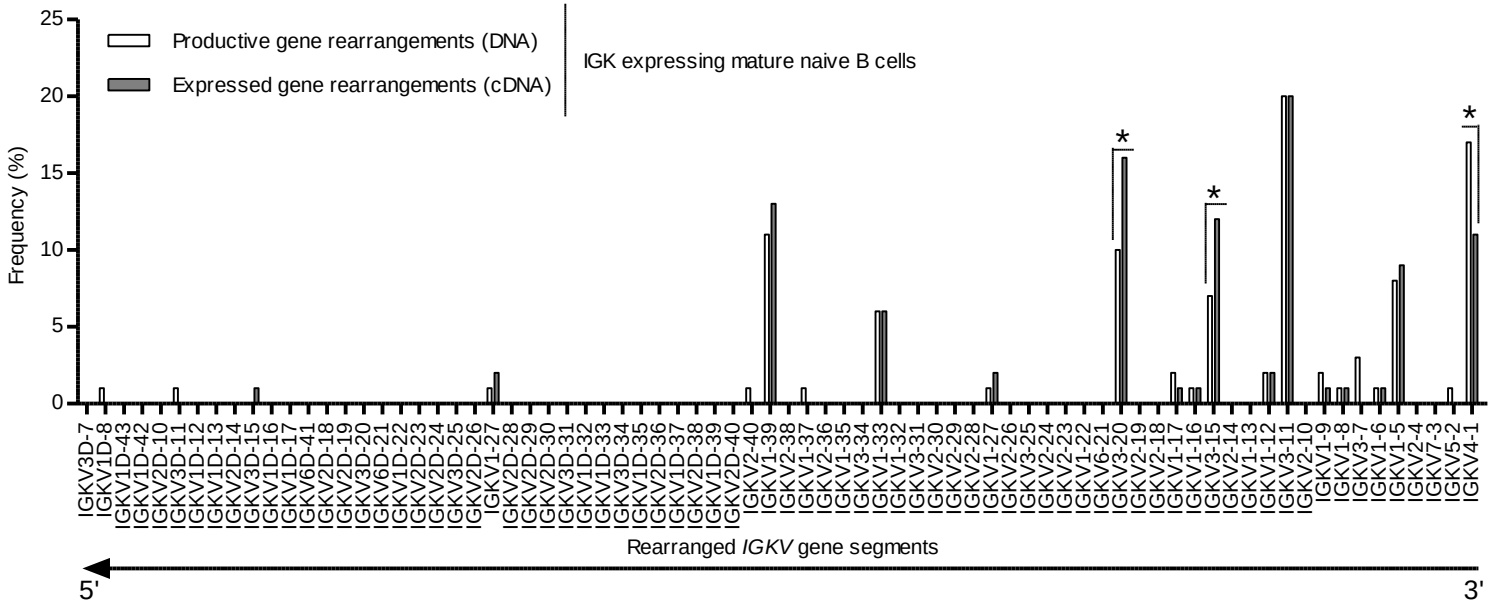


Figure 4-7: Distribution of *IGKV* gene segments in productive and expressed gene rearrangements. Comparison of relative rearrangement frequencies of *IGKV* gene segments in the productively rearranged and expressed cDNA of mature naïve B cells expressing IGK. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). The gene segments *IGKV*3-20 and *IGKV*3-15 were significantly more in the expressed repertoire. Usage of gene segment *IGKV*4-1 was significantly decreased in the expressed repertoire. The gene segments are arranged in order of their position on chromosome 2p11.2.

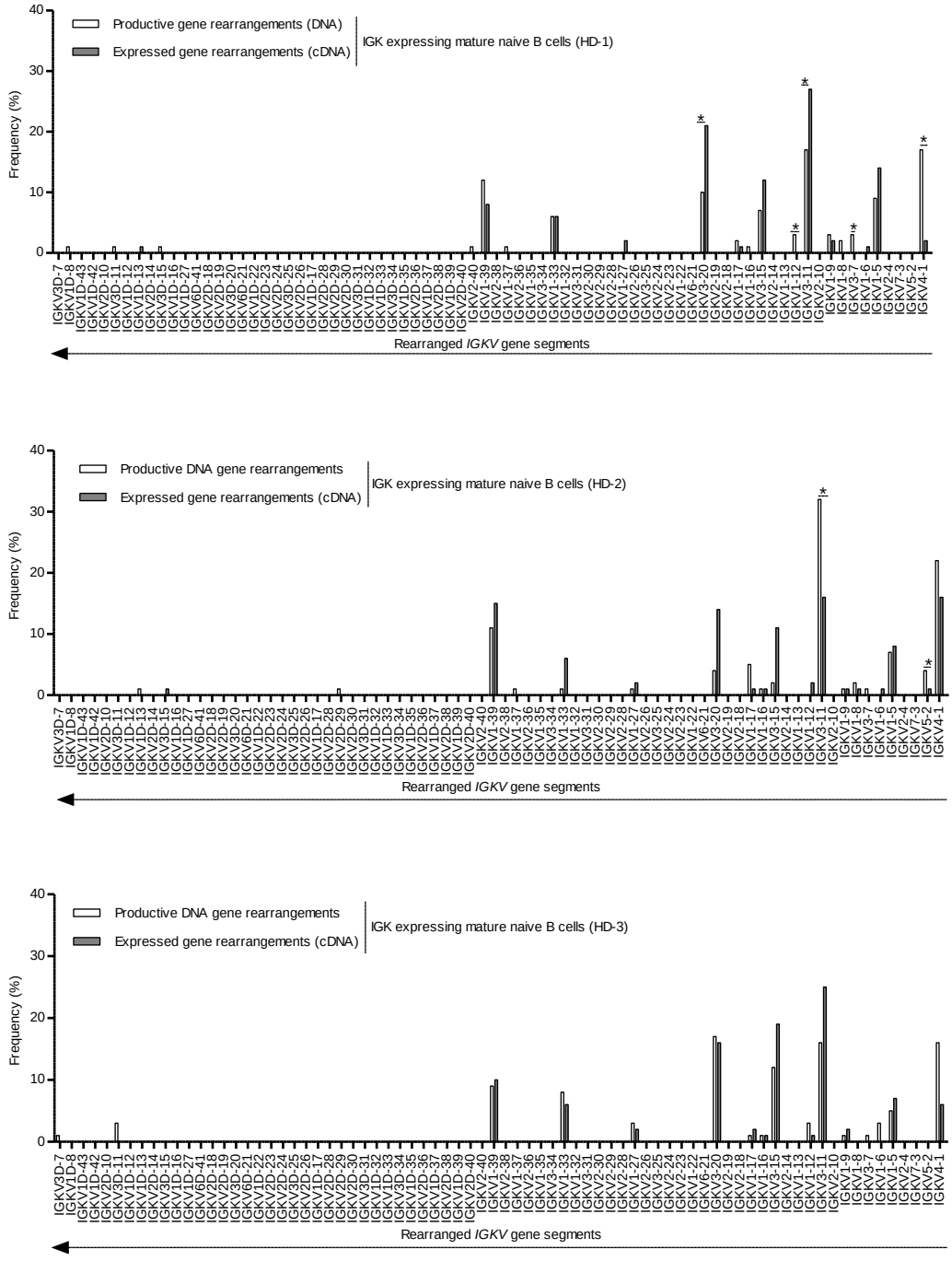


Figure 4-8: Analysis of *IGKV* segments usage. Comparison of relative rearrangement frequencies of *IGKV* gene segments in the productively rearranged and expressed DNA of mature naïve B cells expressing IGK in three healthy donors (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). The gene segments *IGKV*3-11, *IGKV*3-20 and *IGKV*3-15 were significantly more in the expressed repertoire. Usage of gene segment *IGKV*4-1 was significantly decreased in the expressed repertoire. The gene segments are arranged in order of their position on chromosome 2p11.2.

4.2.2 Does the expressed *IGKJ* gene repertoire reflect the productive DNA gene rearrangements?

The differences observed in *IGKV* segment frequency between DNA and cDNA is consistent with receptor editing and accumulation of multiple gene rearrangements that are not expressed. Receptor editing can be investigated by analysis of *IGKJ* segments usage. However, it must be borne in mind those loci where segments (*IGKV* and *IGKJ*) are in the same orientation, editing may not be reflected by representation of downstream *IGKJ* gene segments since segments are not always deleted. The *IGK* locus includes genes in both forward and reverse orientations and gene rearrangements with reverse transcriptional orientation can be retained on the same chromosome. These rearrangements can simultaneously be inactivated by the KDE. On the other hand, gene rearrangements involving *IGKV* gene segments having forward orientation result in the deletion of intervening DNA as an episome that is retained by the cell. Usage of *IGKJ* genes was compared between productive DNA rearrangements with the expressed DNA gene rearrangements. It was found in the genomic repertoire that there was a bias towards more usage of *IGKJ*2 and *IGKJ*5 gene segments and in the expressed repertoire there was significantly more usage of *IGKJ*4 and *IGKJ*2 gene segments (Figure 4-9).

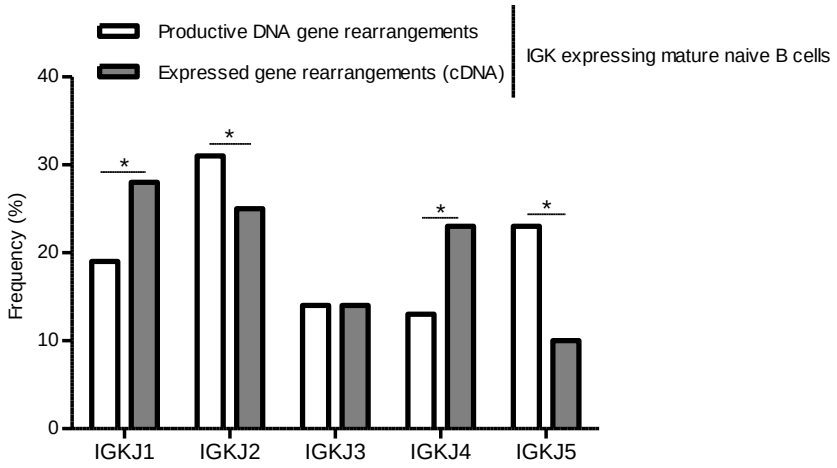


Figure 4-9: Analysis of *IGKJ* segments usage. Comparison of relative rearrangement frequencies of *IGKJ* gene segments in the productive and expressed gene rearrangements of mature naïve B cells expressing IGK. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGKJ*1 and *IGKJ*4 were significantly used more in the expressed repertoire (cDNA). However, *IGKJ*2 and *IGKJ*5 were used more in the productive genomic repertoire (DNA).

The sequences from each individual were then analyzed separately. It was found that all three individuals showed the trend of more usage of *IGKJ*1 and *IGKJ*4 while reduction in the usage of *IGKJ*5 in the expressed repertoire. However, decrease in the usage of *IGKJ*2 in the expressed as observed in the pooled data was not consistent in all donors (Figure 4-10).

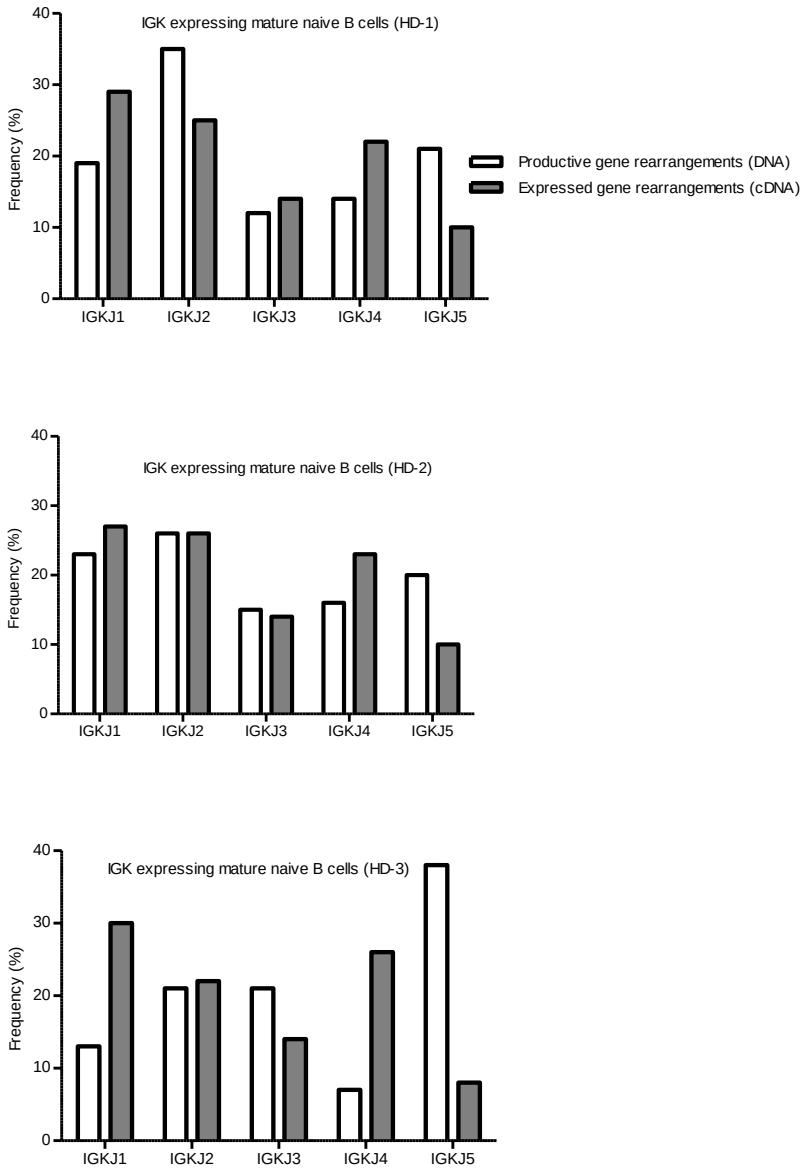


Figure 4-10: Analysis of *IGKJ* segments usage. Comparison of relative rearrangement frequencies of *IGKJ* gene segments in the productive and expressed gene rearrangements of mature naïve B cells expressing IGK in three healthy donors (HD-1 to HD-3). Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). The gene families *IGKJ*1 and *IGKJ*4 were used more in the expressed repertoire (cDNA). However, *IGKJ*5 family were used more in the productive genomic repertoire (DNA).

Data in Figures 4-7 and 4-8 show that gene segments of *IGKV*3 family were represented significantly more frequently in the expressed repertoire as compared to the productive gene

rearrangements and vice versa for *IGKV*4. It was therefore considered possible that different *IGKV* genes may be edited differently and this might be reflected in differences of *IGKJ* rearranged to different *IGKV* gene segments. Therefore, the relative usage of *IGKJ* families was compared in individual gene segments *IGKV*3-11, *IGKV*3-15, *IGKV*3-20, *IGKV*1-39 and *IGKV*4-1. The rationale for selecting these gene segments has been summarized in Table 4-4 based on Figure 4-7.

A similar trend was found in the distribution of *IGKJ*4 and *IGKJ*5 in all *IGKV* gene segments. *IGKJ*5 was used more in the productive gene rearrangements while *IGKJ*4 was used more in the expressed repertoire. However, in *IGKV*1-39 there was no difference in the usage of *IGKJ*5 (Figure 4-11).

Table 4-4: Frequency distribution of selected *IGKV* gene segments in DNA and cDNA

***IGKV* gene segment**	**Features of usage in DNA and cDNA**
*IGKV*1-39	Commonly rearranged but there was no difference in the rearrangement frequency between DNA and cDNA.
*IGKV*3-11	Commonly rearranged but there was no difference in the rearrangement frequency between DNA and cDNA.
*IGKV*3-15	Commonly rearranged and the gene rearrangements were more abundant in cDNA than DNA.
*IGKV*3-20	Commonly rearranged and the gene rearrangements were more abundant in cDNA than DNA.
*IGKV*4-1	Commonly rearranged and the gene rearrangements were more abundant in DNA than cDNA.

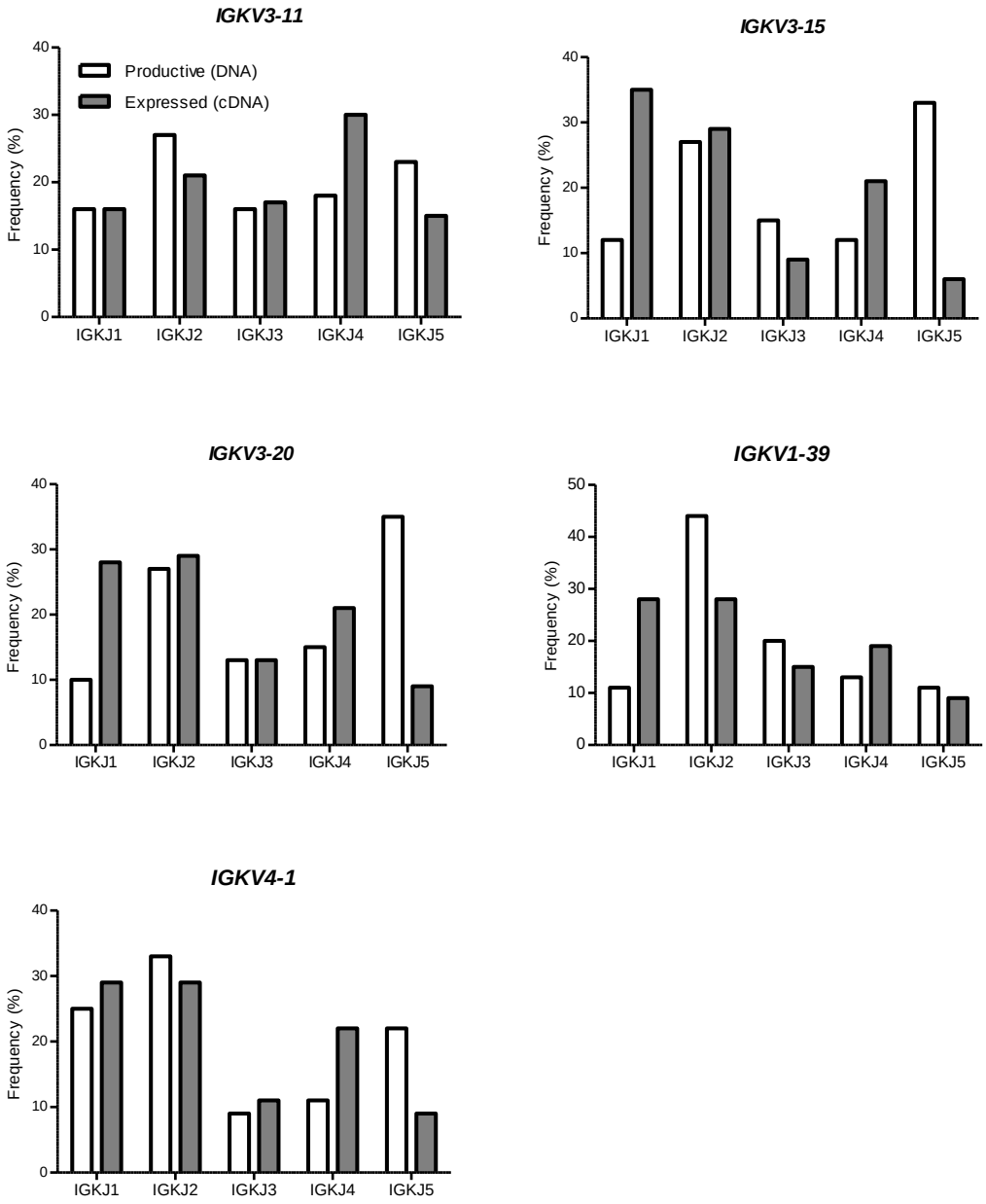


Figure 4-11: Analysis of *IGKJ* segments usage during rearrangements involving *IGKV*3-11, *IGKV*3-15, *IGKV*3-20, *IGKV*1-39 and *IGKV*4-1. Comparison of relative rearrangement frequencies of *IGKJ* gene segments in the productive and expressed gene rearrangements of mature naïve B cells expressing IGK during rearrangement involving selected *IGKV* gene segments. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). The gene families *IGKJ*1 and *IGKJ*4 were used more in the expressed repertoire. However, *IGKJ*5 family was used more in the productive genomic repertoire (DNA) as compared to the expressed repertoire (cDNA).

4.2.3 What are the inherent biases during DNA gene rearrangement process?

The differences between the frequencies of productively rearranged and expressed gene segments suggest that editing occurs during the development of functional *IGK* repertoire resulting in the accumulation of unused productive gene rearrangements. The profile is complex however due to forward and reverse transcription orientations of the *IGKV* gene segments relative to *IGKJ* gene segments on the chromosome. To understand editing and segment selection further, comparison to other groups of sequences was made.

As non productive gene rearrangements are not translated into protein, their presence does not influence the fate of the B cells expressing them. Hence, analyses of non productive gene rearrangements permit an assessment of the inherent biases in rearrangement that contribute to the composition of repertoire. If this assumption is correct, then both sets of independently derived non productive rearrangements of *IGK* from IGK or IGL expressing B cells should resemble each other. Consistent with this, frequencies of gene rearrangements of *IGKV* gene segments were similar in the cells irrespective of surface light chains. The only exception was *IGKV*1-5 which was found to be significantly more in IGK expressing B cells.

It was found that *IGKV* gene rearrangements in the proximal cluster tend to rearrange more frequently. The most common rearranged gene segments in order of their abundance were found to be *IGKV*4-1, *IGKV*3-11, *IGKV*1-39, *IGKV*3-20 *IGKV*1-33, *IGKV*1-5, *IGKV*5-2, *IGKV*3-7, *IGKV*1-37, and *IGKV*3D-11. *IGKV*3D-11 was the only rearranged gene from the distal cluster (Figure 4-12).

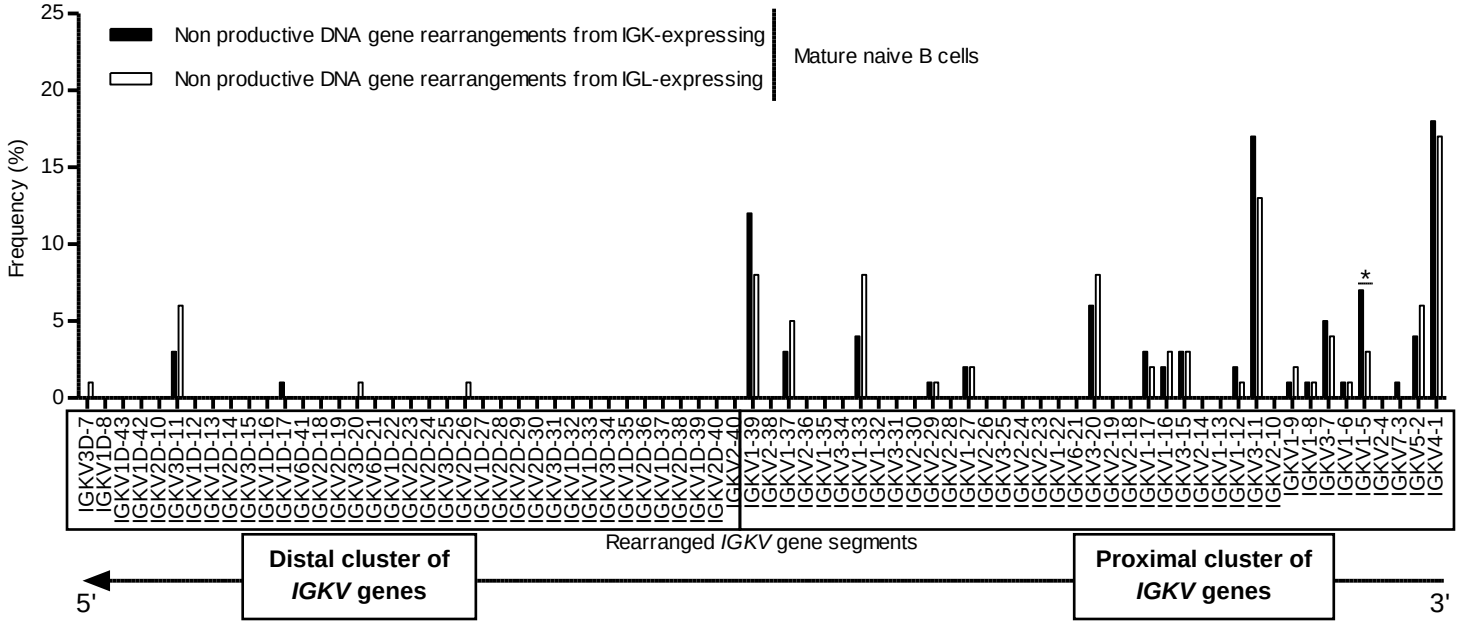


Figure 4-12: Analysis of *IGKV* segments usage. Comparison of relative rearrangement frequencies of *IGKV* gene segments in the non-productively rearranged DNA (*IGKV-IGKJ*) of mature naïve B cells expressing either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). It was found that gene segment *IGKV*1-5 was significantly rearranged more in the IGK expressing B cells. All other *IGKV* gene segments were found to be rearranged at similar frequencies. The gene segments are arranged in order of their position on chromosome 2p11.2.

4.2.4 Is there any selection for or against *IGKV* gene segments?

The decision to rearrange the *IGL* locus is generally considered to be made once the gene rearrangement at the *IGK* locus is unsuccessful on both maternal and paternal alleles. However, productive gene rearrangements of *IGK* locus have been identified in IGL expressing B cells in this study (Figure 4-5) and previously, therefore it was considered that aberrant productive rearrangements of *IGK* derived from IGL expressing B cells have been selected against during B cell development.

The relative usage of *IGKV* families and *IGKV* gene segments was compared between productive gene rearrangements from IGK expressing naïve cells with aberrant productive gene rearrangements from IGL expressing naïve B cells. It was found that there were significantly more rearrangements involving *IGKV*5 gene family in the aberrant productive rearrangements. However, there was no difference in the frequency of rearrangements of other *IGKV* gene families (Figure 4-13).

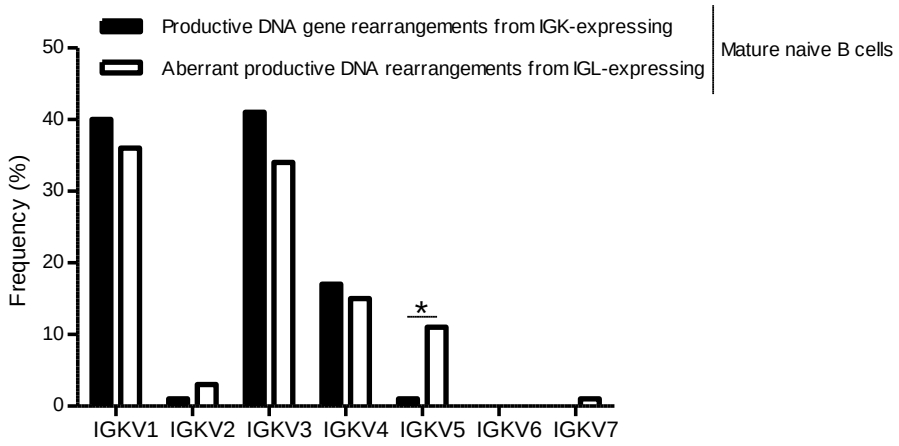


Figure 4-13: Distribution of *IGKV* gene families. Comparison of relative rearrangement frequencies of *IGKV* gene families between productive and aberrant productive DNA gene rearrangements of mature naïve B cells expressing either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Significantly more productive rearrangements utilizing *IGKV*5 family were found in IGL expressing mature naïve B cells.

In terms of gene rearrangement frequencies of *IGKV* gene segments it was found that three *IGKV* gene segments, *IGKV*3D-11, *IGKV*1-37 and *IGKV*5-2 were significantly more frequent in the aberrant productive rearrangements suggesting that they were selected against (Figure 4-14). Productive rearrangements of *IGKV* gene segments such as *IGKV*1-39, *IGKV*3-15, *IGKV*3-11 and *IGKV*1-5 were significantly more frequent in IGK expressing B cells. This could imply that there is bias in the favour of these *IGKV* gene segments (selected for).

To investigate this further, the frequencies of these *IGKV* gene segments in the non productive rearrangements, productive, aberrant productive and used sequences in cDNA were compared (Figure 4-15). Also included was *IGKV*3-7 that is known to be defective due to a non functional splice acceptor site and *IGKV*4-1 that appeared to be selected against by comparison of cDNA and productive gene rearrangements in DNA. It was found that gene rearrangements involving segments *IGKV*3D-11, *IGKV*1-37 and *IGKV*5-2 were more in the productive gene rearrangements of IGL expressing B cells than IGK expressing B cells. Furthermore, the gene rearrangements involving these segments were not found in the cDNA. However, the frequencies of gene rearrangements involving *IGKV*4-1 were similar in the IGK and IGL expressing B cells but cDNA was significantly reduced.

On the other hand, gene segments that were selected for (*IGKV*1-39, *IGKV*3-15, *IGKV*3-11 and *IGKV*1-5) were more in the productive gene rearrangements and cDNA of IGK expressing B cells (Figure 4-16).

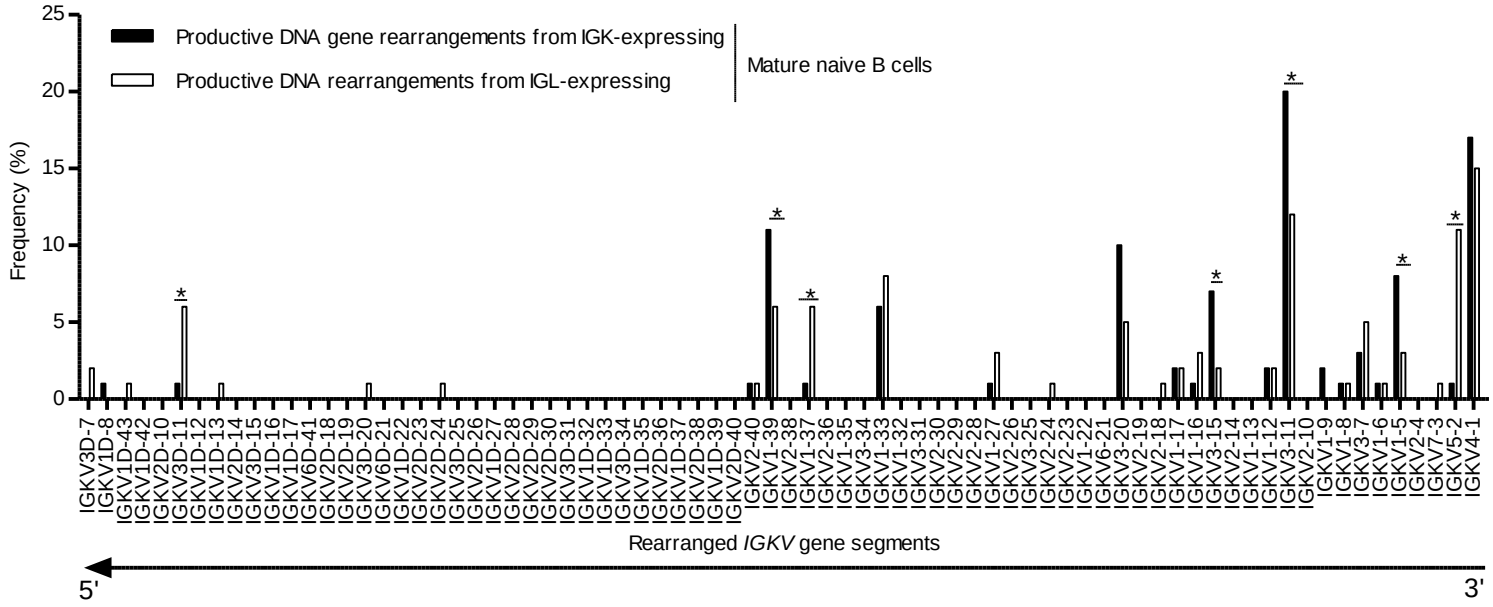


Figure 4-14: Analysis of *IGKV* segments usage. Comparison of relative rearrangement frequencies of *IGKV* gene segments in the productively rearranged DNA of mature naïve B cells expressing either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Gene segment *IGKV*1-39, *IGKV*3-15, *IGKV*3-11, were significantly more productively rearranged in IGK expressing mature naïve B cells. Gene segments *IGKV*3D-11, *IGKV*1-37 were significantly rearranged more productively in the DNA of IGL expressing mature naïve B cells. The gene segments are arranged in order of their position on chromosome 2p11.2.

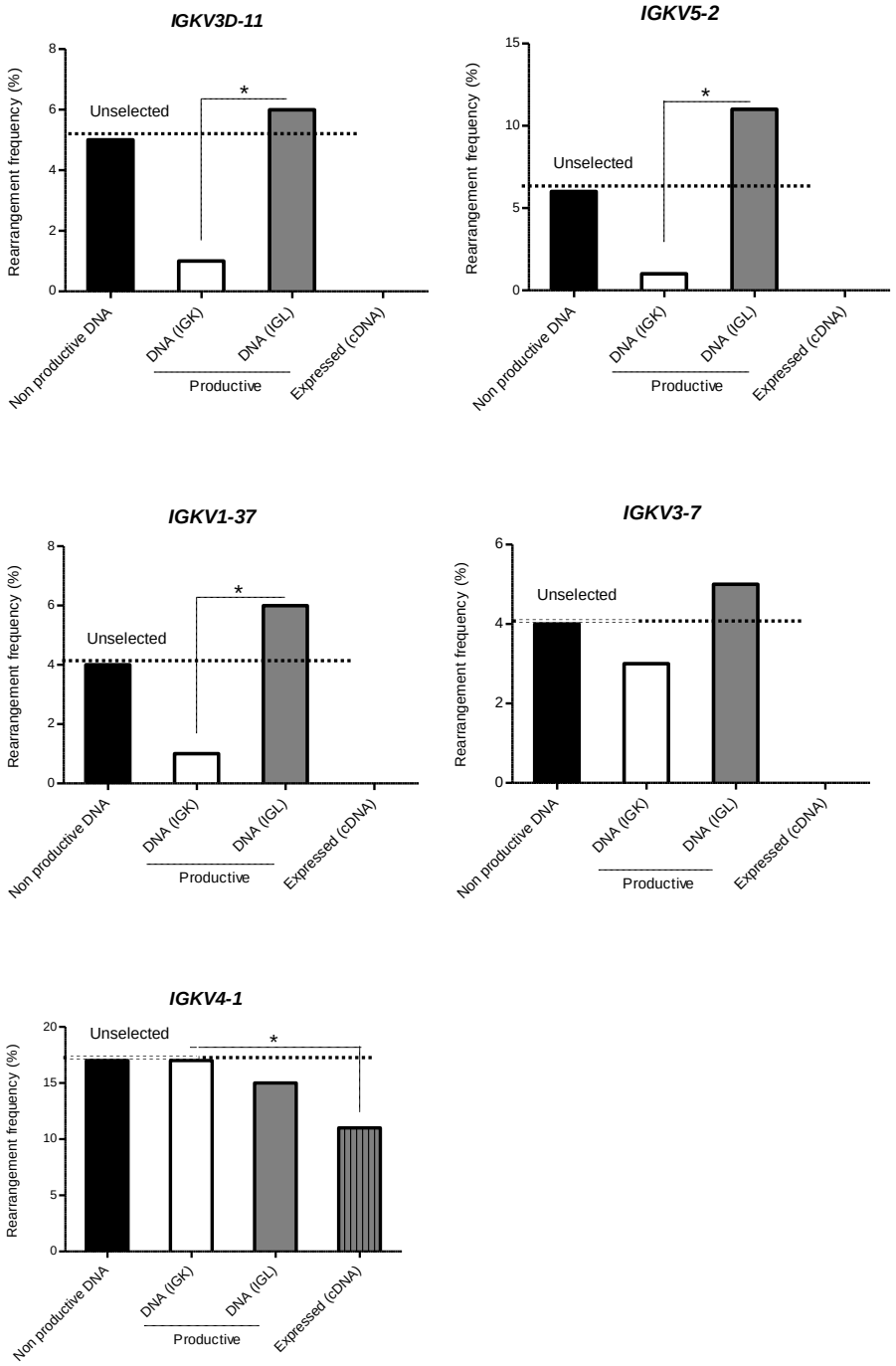


Figure 4-15: Analysis of *IGKV* gene segments (*IGKV*3D-11, *IGKV*5-2, *IGKV*1-37, *IGKV*3-7 and *IGKV*4-1) selected against during rearrangement at *IGK* locus. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). However, *IGKV*3D-11, *IGKV*5-2, *IGKV*1-37 and *IGKV*3-7 were more in the productive rearrangements of IGL expressing B cells. Furthermore, these rearrangements were not found in the expressed repertoire. There was no difference in the rearrangement frequencies of *IGKV*4-1 between IGK and IGL expressing B cells but these were significantly less abundant in the cDNA.

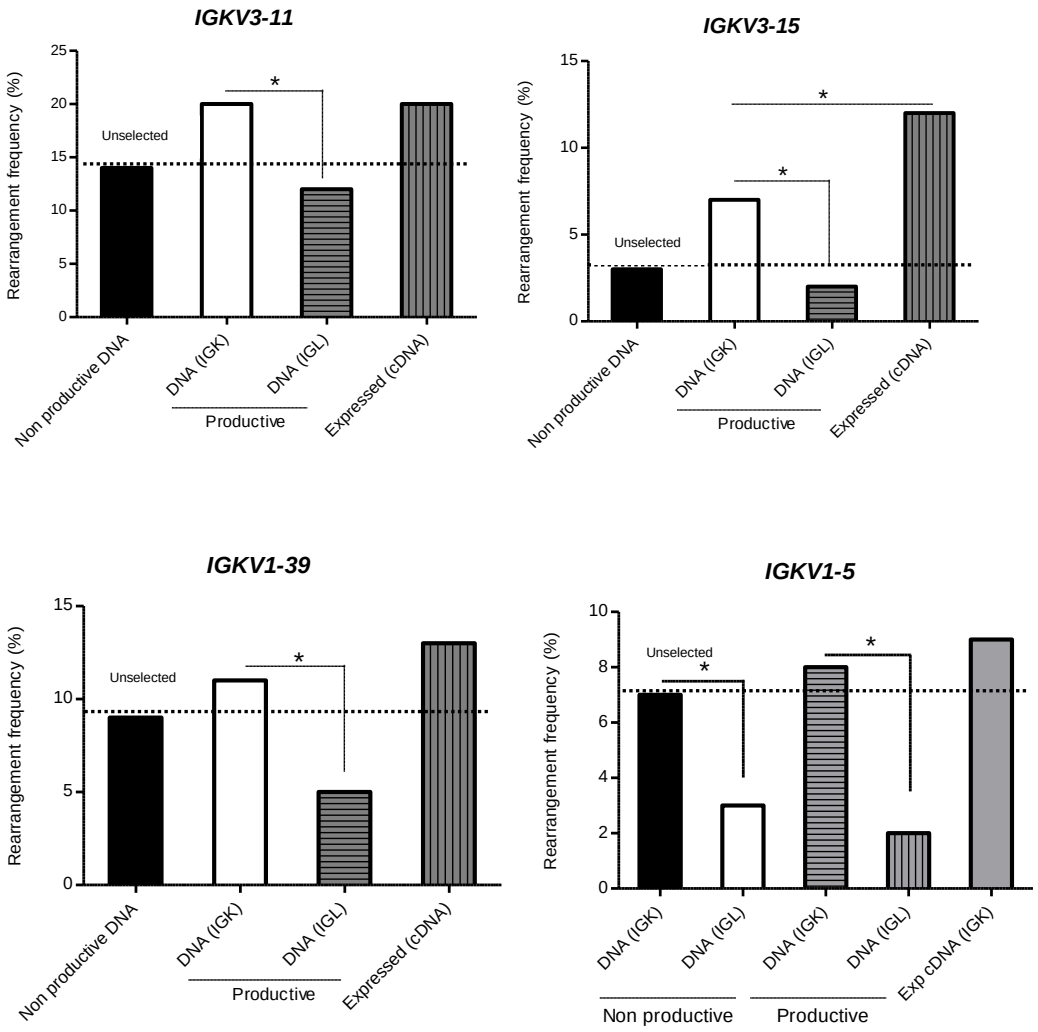


Figure 4-16: Analysis of *IGKV* gene segments (*IGKV3*-11, *IGKV*3-15, *IGKV*1-39, and *IGKV*1-5) selected during rearrangement at *IGK* locus. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Frequencies of non productive gene rearrangements indicate all segments were frequently rearranged. However, there were significantly more productive rearrangements in IGK expressing than IGL expressing B cells. Furthermore, these rearrangements were also found in the expressed repertoire.

4.2.5 Is there any difference in the characteristic of CDR-3 region of aberrant productive, productive and expressed repertoire?

The CDR-3 region encompasses the junction of *IGKV-IGKJ* region. It contributes to determination of the antigenic specificity. It has been found that after antigenic selection shorter CDR-3 is selected. Longer CDR-3s have been found to be associated with DNA binding autoantibodies. In this section CDR-3 characteristics of the non productive, productive, aberrant productive and expressed DNA gene rearrangements were compared to see if selection is apparent in this data. It was hypothesized that there might be a difference in the CDR-3 characteristics of the aberrant productive rearrangements that could potentially affect antigen recognition or subsequent pairing with IGH negatively. The following features of CDR-3 were compared:

1. CDR-3 length
2. Relative distribution of number of amino acids
3. GRAVY indices
4. Isoelectric points
5. Aliphatic indices

4.2.5.1 CDR-3 length (amino acids)

The mean length of CDR-3 region was compared among non productive, aberrant productive, productive and expressed gene rearrangements. The mean length of CDR-3 was found to be 8.4 amino acids, 8.9 amino acids, 9.3 amino acids and 9.2 amino acids in the non productive, aberrant productive, productive and expressed gene rearrangements (Figure 4-17). The CDR-3 length of non productive and aberrant productive rearrangements was significantly shorter than productive and expressed gene rearrangements.

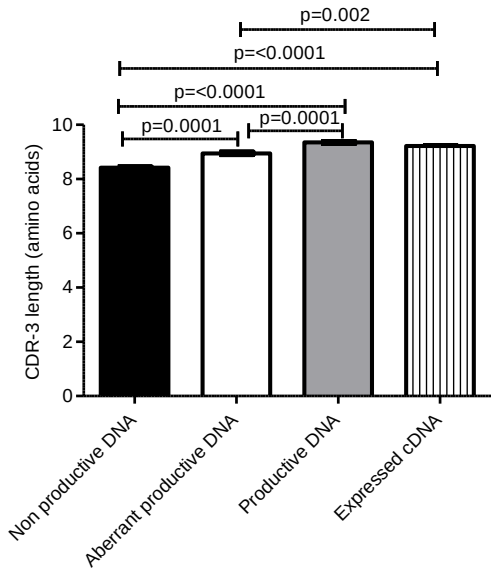


Figure 4-17: The mean CDR-3 length of non productive, aberrant productive, productive and expressed DNA rearrangements. Non parametric tests were used to compare the means. The CDR-3 length of non productive and aberrant productive gene rearrangements was significantly shorter than productive and expressed gene rearrangements.

4.2.5.2 Relative distribution of number of amino acids

The physicochemical properties of proteins are affected by the number of amino acids in the CDR-3 region. Therefore the relative distribution of the number of amino acids in different sets of gene rearrangements was compared. It was found in the aberrant productive DNA gene rearrangements that there were some differences in the relative proportion of different number of amino acids. The mean CDR-3 length was found to be 9 amino acids in the expressed repertoire suggesting the selection of CDR-3 of 9 amino acids length but in the

aberrant productive rearrangements there were more sequences that were larger or shorter than this (Figure 4-18).

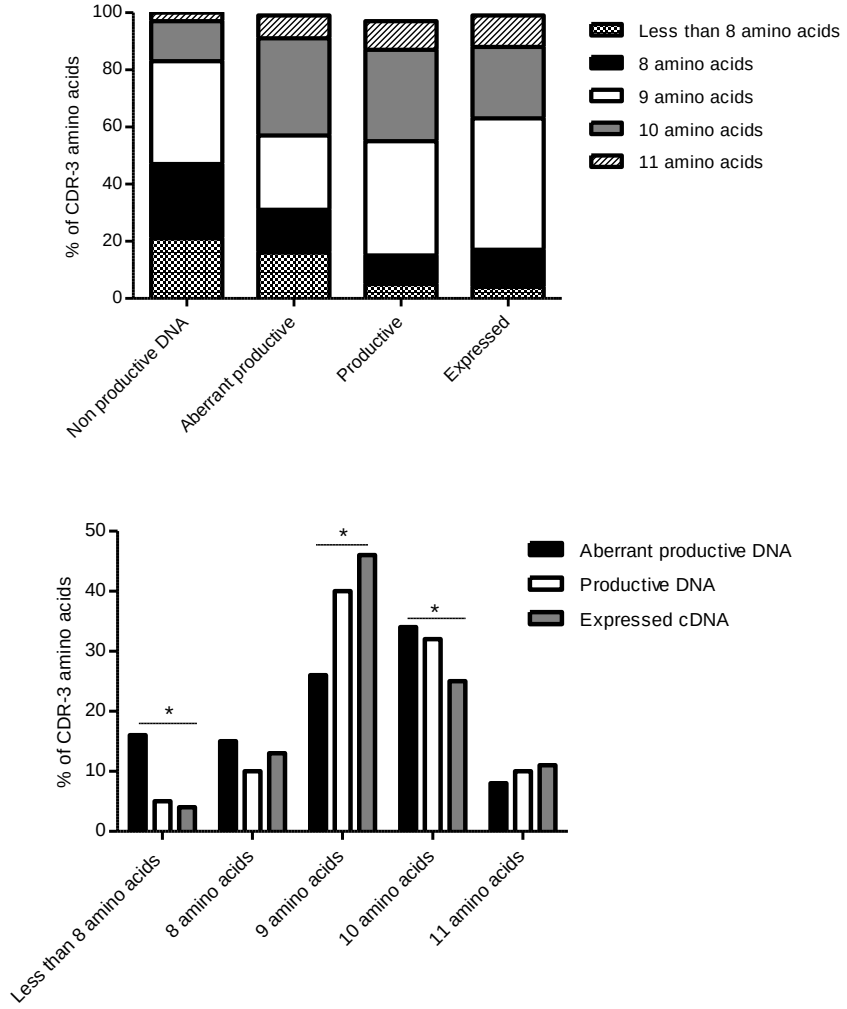


Figure 4-18: The relative distribution of amino acids in the non productive, aberrant productive, productive and expressed DNA gene rearrangements.

There was a difference in the relative proportion of CDR-3 amino acids in the expressed and aberrant productive rearrangements. Therefore, ensuing CDR-3 characteristics were compared according to the number of amino acids (8, 9, 10, 11 amino acids) because these characteristics contribute to the physicochemical properties of the proteins.

4.2.5.3 GRAVY indices

GRAVY index indicates the solubility of the proteins. Positive values indicate a protein is hydrophobic while negative values indicate a hydrophilic protein. The GRAVY index is calculated by adding the hydropathy value for each residue and dividing by the length of the sequence (Kyte and Doolittle 1982). Due to variation in the proportion of amino acids in the aberrant productive, productive and expressed repertoire GRAVY indices were compared for each group (8, 9, 10,11 amino acids). In the group of 8, 9 and 10 amino acids there was no difference in the GRAVY indices. However, in the group of CDR-3 consisting of 11 amino acids the GRAVY index (more hydrophilic) of the productive and expressed repertoire was significantly lower (Figure 4-19).

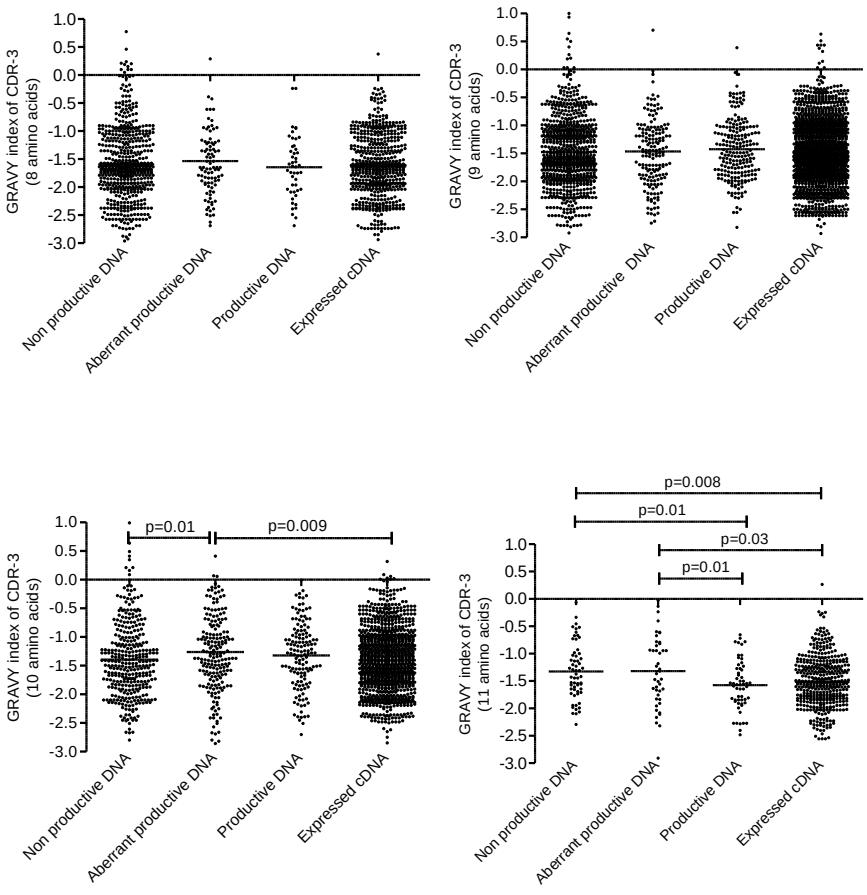


Figure 4-19: The GRAVY indices of CDR-3 regions having 8, 9 10 and 11 amino acids. ProtPram was used to determine the aliphatic index of the CDR-3 region and the grand average hydropathy (GRAVY) index. Non parametric tests were used to compare the means. Significant differences were observed for the CDR-3 having 11 amino acids.

4.2.5.4 Isoelectric point

The isoelectric point is the pH at which a particular molecule or surface carries no net electrical charge. The net charge on the molecule is affected by the pH of its surrounding environment and can become more positively or negatively charged due to the gain or loss, of protons respectively. Immunoglobulins vary in their isoelectric points ranging within pH 4 to 9. Isoelectric points were compared among aberrant productive, productive and expressed repertoire for each group of amino acids in CDR-3. No significant differences were observed in any group of amino acids (Figure 4-20).

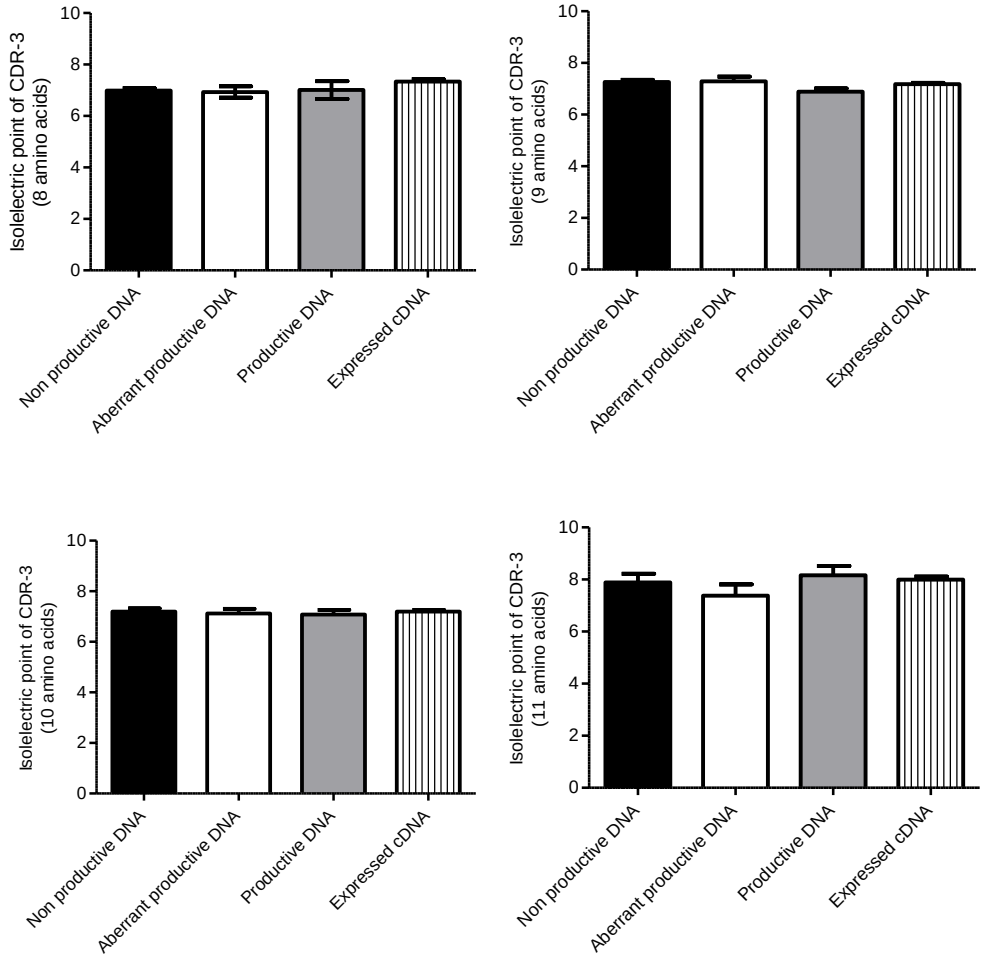


Figure 4-20: The isoelectric point of CDR-3 regions having 8, 9 10 and 11 amino acids. Non parametric tests were used to compare the means. No significant differences were observed for the different CDR-3 lengths.

4.2.5.5 Aliphatic indices

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins. Aliphatic indices were compared for each group of CDR-3 amino acid lengths (8, 9, 10 and 11). It was found that irrespective of the CDR-3 length, aliphatic indices of productive and expressed repertoire were less than the aberrant productive group of rearrangements (Figure 4-21).

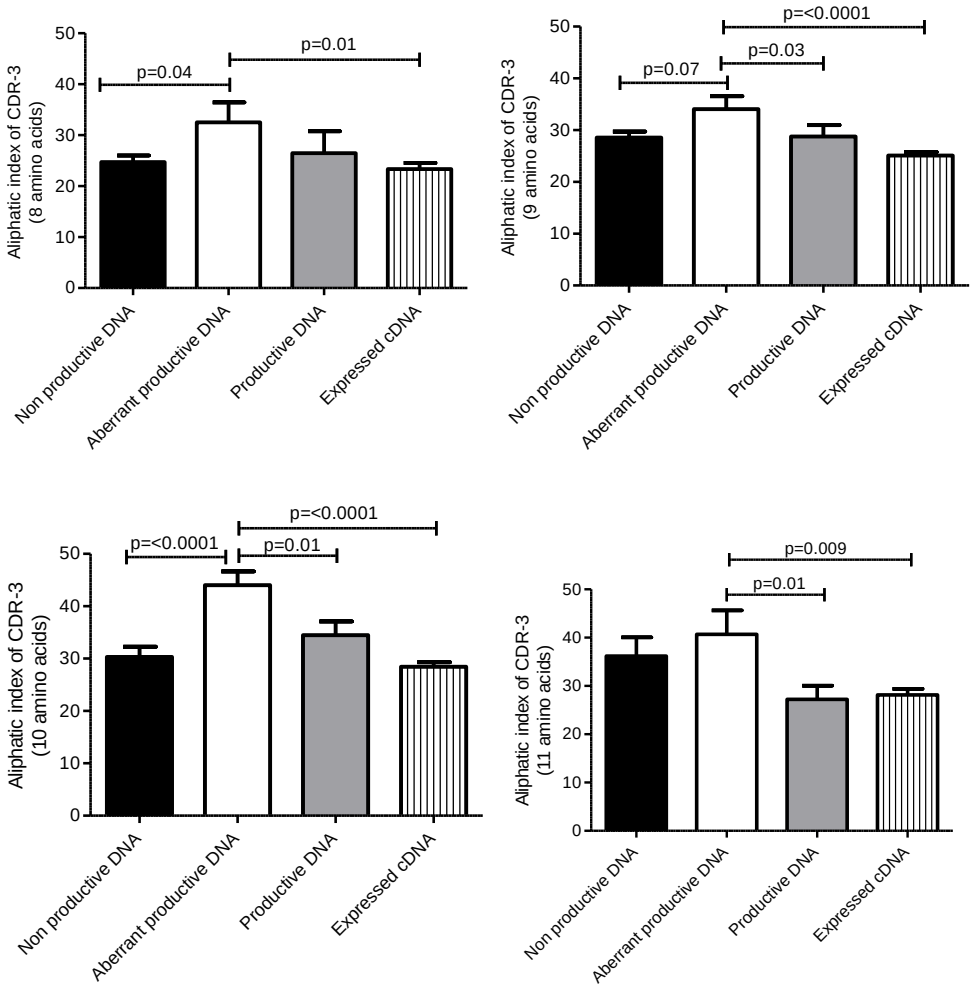


Figure 4-21: The aliphatic indices of CDR-3 regions having 8, 9, 10 and 11 amino acids. Non parametric tests were used to compare the means. Significant differences were observed irrespective of CDR-3 lengths. Aberrant productive gene rearrangements have a higher aliphatic index.

4.3 Discussion

4.3.1 The rearrangement frequency of *IGKV4-1* differs between DNA and the expressed repertoire

In this chapter, productive rearrangements of *IGK* were compared between DNA and the expressed cDNA in the same sorted population of IGK expressing mature naïve B cells. It was found that gene segment *IGKV*4-1 was abundantly rearranged in the DNA but significantly reduced in the expressed repertoire and this segment was the only with these properties. This was initially considered to be surprising since productive rearrangements encode the expressed repertoire. The possibility that this could be an artefact due to PCR protocols used for DNA and cDNA was considered. It was thought unlikely however, since the same set of variable region primers were used for both PCR reactions.

The gene segment *IGKV*4-1 was found to be consistently abundantly rearranged in the DNA of all individuals. Some molecular and topological features may tend to favour the relatively high rearrangement frequency of *IGKV*4-1 in the DNA. The gene segment *IGKV*4-1 is present in the inverted orientation and is in very close proximity to the *IGKJ* gene segments (Foster, Brezinschek et al. 1997). Germline transcripts of *IGKV*4-1 have been found in pre B cells and fetal bone marrow cells implying that this segment is readily available for recombination (Martin, Huang et al. 1991).

Previously gene segment *IGKV*4-1 has been reported to be over represented in the B cells that were in the process of receptor editing (Meffre, Davis et al. 2000). The ligated receptor may stimulate further gene rearrangements that replace previous receptor genes on the same allele. Several secondary gene rearrangements are possible on the same allele unless the cell gets exhausted after the deletion of the most downstream *IGKJ*5 gene segment (Bergman and Cedar 2004).

The gene segment *IGKV*4-1 has also been implicated in the expressed repertoire of some autoimmune diseases such as diabetes and SLE (Dorner, Foster et al. 1998, Woodward and Thomas 2005). Therefore, it is quite possible that *IGKV*4-1 is abundantly rearranged due to its accessibility and proximity to the *IGKJ* segments but it might not pass selection criteria and this can lead to the receptor editing.

The inverted transcriptional orientation of *IGKV*4-1 could favour the accumulation of the rearrangements as inverted DNA on the same chromosome. For example if the primary productive gene rearrangement takes place between *IGKV*4-1 and *IGKJ*2 gene segments then it would result in the inversion of intervening DNA (as depicted by the inversion of *IGKJ1* in Figure 4-22). If for some reason, primary potentially productive *IGKV*4-1-*IGKJ*2 rearrangement is selected against then there could be two further possibilities:

1. The downstream *IGKJ* gene segments (*IGKJ*3, *IGKJ*4 or *IGKJ*5) can rearrange with any upstream *IGKV* gene segment of proximal cluster. It will result in the deletion of intervening DNA including productive joint *IGKV*4-1-*IGKJ*2 rearrangement. The deleted part is retained as an episome. Rearranged *IGKV*4-1-*IGKJ*2 could be amplified from the naïve B cells (Figure 4-22A).

2. An inverted gene segment *IGKJ*1 still retains the capability to rearrange with any upstream *IGKV* gene segment by inversion. In this way the secondary gene rearrangement and primary selected against *IGK*4-1-*IGKJ*2 rearrangement can be retained on the same chromosome (Figure 4-22B). Consequently, 2 rearrangements could be amplified by PCR from the same chromosome.

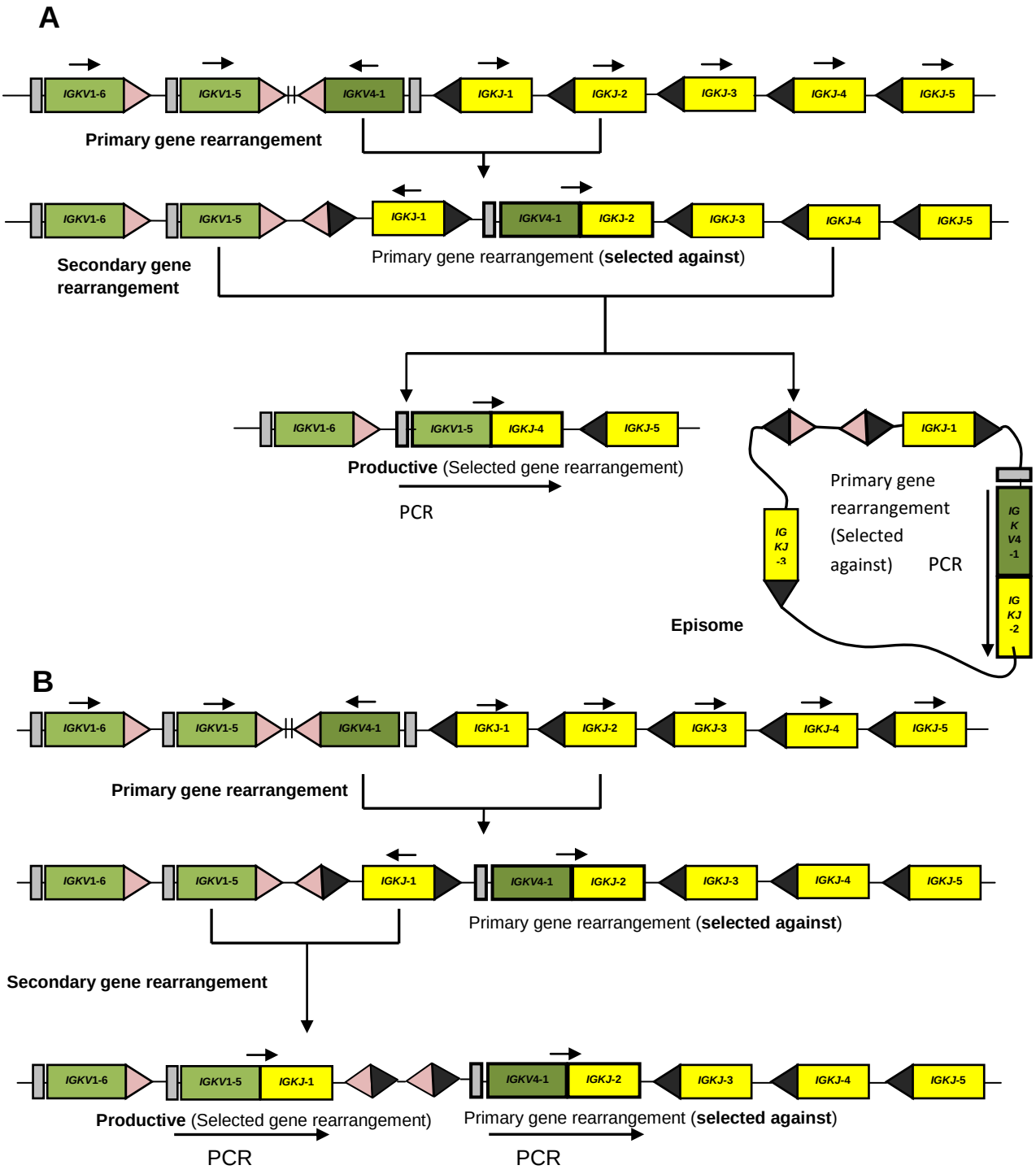


Figure 4-22: Theoretical accumulation of gene rearrangements of *IGKV*4-1 that have been selected against by editing. The arrows indicate the transcriptional orientation of various gene segments. **(A)** The primary non selected gene rearrangement between *IGKV*4-1 and *IGKJ*2 may be followed by the secondary gene rearrangement of an upstream *IGKV*1-5 and downstream *IGKJ*4. This results in the deletion of intervening DNA but retained as an episome within a cell. **(B)** The primary gene rearrangement between *IGKV*4-1 and *IGKJ*2 may be followed by the secondary gene rearrangement of upstream *IGKV*1-5 and inverted *IGKJ*1 gene segment. This results in the retention of secondary gene rearrangement between *IGKV*1-***5*** and *IGKJ*1. Various productive and unselected productive gene rearrangements can be amplified from one chromosome and not distinguished by PCR.

In either case multiple selected against productive and productive coding joints would remain inside the cell and could potentially be detected by PCR. After all possible gene rearrangements on one allele, the process of gene rearrangements can be initiated on the second allele. Rearrangements of *IGKV*4-1 could also accumulate in a cell if they were not favoured when rearranged on the first allele so that the cell moves on to rearrange the second allele.

Overall and despite the mechanism, higher frequency of *IGKV*4-1 in the DNA versus cDNA suggests selection against and use of this segment.

4.3.2 The gene segments *IGKJ*5 is abundantly rearranged in the DNA but *IGKJ*4 is predominant in the expressed repertoire

There were some differences in the relative usage of *IGKJ* gene segments between the productive and expressed repertoire. It was found in the productive gene rearrangements that there was a significant bias towards usage of *IGKJ*5 while the expressed repertoire was skewed towards *IGKJ*4 usage. This trend was consistent among all individuals. The same pattern was found when usage of *IGKJ* genes was compared in different *IGKV* families between DNA and cDNA. The most commonly used *IGKJ* segments were *IGKJ*1 and *IGKJ*2 followed by *IGKJ*4 and *IGKJ*5 respectively. The least used segment was *IGKJ*3. This pattern of *IGKJ* usage in the expressed repertoire is consistent with previous studies (Gay, Saunders et al. 1993, Weber, Blaison et al. 1994, Collins, Wang et al. 2008).

Conventionally usage of upstream *IGKJ* gene segments suggests receptor editing (Foster, Brezinschek et al. 1997). However, this might not be reflected at the *IGK* locus. The *IGK* locus includes genes in both forward and reverse orientations and gene rearrangements with reverse transcriptional orientation can be retained on the same chromosome (Zheng, Wilson et al. 2004). These rearrangements can simultaneously be inactivated by the KDE (Graninger, Goldman et al. 1988, Vela, Ait-Azzouzene et al. 2008). On the other hand, gene rearrangements involving *IGKV* gene segments having forward orientation result in the deletion of intervening DNA as an episome that is retained by the cell.

There could be a technical reason that can generate bias in *IGKJ* usage. For the amplification of genomic and cDNA, the same *IGKV* primers were used for the 5’ end.

However, for the 3' end *IGKJ* and *IGKC* primers were used for DNA and cDNA respectively. This could affect *IGKJ* usage but not *IGKV*.

4.3.3 There are some inherent biases in the tendency of segments to rearrange at *IGK* locus

Non productive gene rearrangements are not influenced by any preimmune selection pressure and can predict base level of gene rearrangements. There was no difference in the non productive gene rearrangements at *IGK* locus irrespective of the surface BCR except relative usage of *IGKV*1-5. The gene segment *IGKV*1-5 was found more in the rearrangements from the cell expressing IGK chain and it was consistent among all three individuals.

The data indicates that *IGKV* genes in the proximal cluster tend to rearrange more frequently and this is consistent with other studies (Foster, Brezinschek et al. 1997, Jackson, Kidd et al. 2013). The distal cluster is in the inverted orientation and inversional gene rearrangement may be less efficient than the deletional gene rearrangement which may affect the rearrangement frequencies (Gauss and Lieber 1992). Several molecular mechanisms have been reported that may affect non random utilization of *IGKV* genes such as distance of *IGKV* genes from *IGKJ* segments, transcriptional orientation, efficiency of transcription, promoter strength, enhancer activity, spacer length and sequence, RSS efficiency and accessibility of individual segments to recombination machinery (Nadel, Tang et al. 1998, Feeney, Tang et al. 2000). Recently it has been found that accessibility of immunoglobulin genes is determined by the post translational epigenetic modifications of regional histone cores. Acetylation of histones has also been correlated with a bias in *IGK* gene rearrangements (McMurry and Krangel 2000, Spicuglia, Franchini et al. 2006). Furthermore, accessibility to recombination correlates with transcription. The transcription factors STAT5 and E2A have been implicated to increase the accessibility of *IGK* locus (Mandal, Powers et al. 2011). However, the precise relationship between *IGK* recombination and transcription is still unclear (Hamel, Mandal et al. 2014).

4.3.4 Selection pressures shape *IGK* repertoire in mature naïve B cells

Generally gene rearrangement proceeds to *IGL* locus after failure to produce a successful rearrangement at *IGK* locus. The data presented show 25% *IGK* gene rearrangements in IGL expressing B cells were technically productive (aberrant) gene rearrangement. The presence of such non functional productive *IGK* gene rearrangements has been found in humans and mice previously (Retter and Nemazee 1998, Brauninger, Goossens et al. 2001). Since these rearrangements were not used by the cell but instead a decision was made to start rearrangement at *IGL* locus these rearrangements were classified as aberrant productive and selected against.

Gene segments *IGKV*3D-11*, IGKV*1-37*, IGKV*1-27 *IGKV*3-7 and *IGKV*5-2 were selected against during the development of *IGK* repertoire. The rearrangements involving these segments were significantly more frequent in the productive gene rearrangements of IGL expressing B cells than IGK expressing B cells. Two other gene segments, *IGKV*1-27 and *IGKV*3-7 were also more in the productive rearrangements of IGL expressing B cells although this was not statistically significant.

The gene segment *IGKV*1-27 has a defective RSS due to a single base pair substitution at the 4th position of heptamer (Lautner-Rieske, Huber et al. 1992). The gene segment *IGKV*1-37 is structurally defective in that it encodes glycine in place of cysteine at position 88. The gene segment *IGKV*3-7 is defective in acceptor splice site (Pargent, Meindl et al. 1991). This suggests that structural abnormality or mutation in the regulatory elements can render productive rearrangement as non functional or aberrant. However, it is not known which particular factors are responsible for not selecting the gene rearrangements involving *IGKV*5-2 and *IGKV*3D-11. The gene segments which appeared to be selected against were not found in the cDNA which further emphasizes the rejection of these rearrangements during selection.

Gene segments *IGKV*3-11, *IGKV*3-15, *IGKV*3-20*, IGKV*1-39 and *IGKV*1-5 appeared to be selected for during rearrangement. The gene rearrangements involving these segments were significantly more in the productive gene rearrangements derived from IGK expressing B cells than IGL expressing B cells. Over representation of these segments in the productive

and expressed gene rearrangements is consistent with previous studies (Foster, Brezinschek et al. 1997, Collins, Wang et al. 2008, Jackson, Kidd et al. 2013). In addition to this, the comparison of productive and expressed gene rearrangements revealed *IGKV*3-20 and *IGKV*3-15 were more in the cDNA than DNA. When different individuals were analyzed separately all showed the trend of more utilization of *IGKV*3 family in the expressed repertoire than DNA. These findings suggest that biased appearance of these segments may be related to the events after rearrangement at the *IGK* locus and is therefore dependent on the expression of IGK on the cell surface. Subtle differences in the transcription rates that may occur could also play a role in the over representation of these segments in cDNA, but it is not known if this is a feature of the expressed repertoire (Foster, Brezinschek et al. 1997).

4.3.5 Differences in the CDR-3 region

In addition to changes in the repertoire of *IGKV* and *IGKJ* segments usage, there were also some differences in the CDR-3 region in the groups of sequences studied. During rearrangement, the junction of *IGKV* and *IGKJ* is modified due to the activity of exonucleases and transferases. The CDR-3 region is the major determinant of antigen specificity and the shorter CDR-3 lengths have been associated with the selection of memory B cell population after GC reaction. However, longer CDR-3 lengths of IGH have been implicated in some pathological states (Aguilera, Melero et al. 2001, Miqueu, Guillet et al. 2007, Wu, Kipling et al. 2010). The mean CDR-3 length in the productive and expressed gene rearrangements was found to be 9 amino acids (27 nucleotides). Consistent with this, several previous studies have found that the CDR-3 regions of expressed human IGK are usually 27 nucleotides in length (Bridges, Lee et al. 1995, Tomlinson, Cox et al. 1995, Foster, Brezinschek et al. 1997). On the other hand, mean CDR-3 length of non productive and aberrant productive rearrangements was shorter than productive and expressed gene rearrangements significantly. Previously no difference was found between the CDR-3 length of productive and non productive gene rearrangements which is inconsistent with our findings (Foster, Brezinschek et al. 1997).

4.4 Conclusions

Overall, the data in this chapter identifies gene segments that are selected for and against during B cell development in the bone marrow when fate decisions are made whether to express a functionally rearranged *IGK* gene or move on to rearrange the *IGL* locus. Gene segments *IGKV*3D-11*, IGKV*1-37*, IGKV*1-27, *IGKV*3-7 and *IGKV*5-2 are clearly selected against at this point. Gene segment *IGKV*5-2 is selected against, but this has not been reported previously. Gene segments *IGKV*3-11, *IGKV*3-15, *IGKV*3-20*, IGKV*1-39 and *IGKV*1-5 appear to be favoured in the used repertoire. Gene segment *IGKV4-1* is very interesting because it appears to be frequently rearranged but selected against as it is significantly reduced at the time of expression. *IGKV*4-1 has been linked with autoimmune diseases such as SLE where it appears to be over represented, and it is possible that breakage of the mechanisms that normally select this segment could be associated with the pathogenesis of autoimmune diseases. Based on the objectives of this chapter, following are the concluding points:

1. The expressed repertoire of *IGKV* gene repertoire can be different from the productive gene rearrangements in DNA due to retention of unused gene rearrangements.
2. The expressed repertoire of *IGKJ* gene repertoire can be different from the productive gene rearrangements in DNA.
3. The inherent biases at the *IGK* locus are similar in IGK and IGL expressing B cells.
4. Gene segments *IGKV*3D-11*, IGKV*1-37*, IGKV*1-27 *IGKV*3-7 and *IGKV*5-2 are selected against while gene segments *IGKV*3-11, *IGKV*3-15, *IGKV*3-20*, IGKV*1-39 and *IGKV*1-5 are selected for at the *IGK* locus.
5. The CDR-3 characteristics of aberrant productive gene rearrangements are different from expressed gene rearrangements.

Chapter 5

Immunoglobulin heavy chain variable region of mature naïve B cells expressing either kappa or lambda light chain

5.1 Introduction

It is known that there is non-random usage of *IGHV*, *IGHD* and *IGHJ* gene segments in VDJ rearrangements of IGH expressed by human B cells. In the mature naïve B cell repertoire different *IGHV* gene segments are used with frequencies that range approximately from as little as 0.1% to 10% of all rearrangements (Boyd, Gaeta et al. 2010, Glanville, Kuo et al. 2011). Also, some gene segments tend to be used with approximately equivalent frequencies in different individuals, for example gene segment *IGHV*1-46 is apparent in around 3% of rearrangements, *IGHV*3-21 in around 5% and *IGHV*3-49 in around 1% (Boyd, Gaeta et al. 2010, Briney, Willis et al. 2012, Watson and Breden 2012). Biased usage has also been reported for *IGHD* gene segments where *IGHD*3-22 and *IGHD*4-4/11 are the most frequently used segments (Benichou, Glanville et al. 2013). Usage of *IGHJ* segments is also skewed with *IGHJ*4 and *IGHJ*6 being the most commonly used *IGHJ* gene segments (Volpe and Kepler 2008, Larimore, McCormick et al. 2012, Watson and Breden 2012).

Various factors are known to contribute to variability in *IGH* gene segment usage. It has been found in a murine model that *IGHV* gene segments that are proximal to the *IGHD* gene segments are rearranged preferentially (Yancopoulos, Desiderio et al. 1984). It has been reported in the human fetal liver there are some biases in the usage of *IGHV*, *IGHD* and *IGHJ* gene segments depending upon their proximity on the chromosome. For example *IGHV*6-1 gene segment which is close to the *IGHD* gene segments on the chromosome tend to rearrange more frequently. Similarly the *IGHDQ*-52 gene segment which is close to the *IGHJ* gene segments rearranges preferentially (Jung, Giallourakis et al. 2006, Schelonka, Szymanska et al. 2010). Various factors that may affect the preferential rearrangement of *IGHV* gene segments include the RSS strength associated with each segment that is determined by the actual sequence of the RSS and also differential accessibility of individual *IGHV* gene segments that is defined by local transcriptional activity (Jung, Giallourakis et al. 2006). The proximal gene segments are more accessible to the RAG proteins. It has been found that variations in the canonical RSS can decrease the efficiency of gene rearrangement that in turn decreases the frequency at which a gene segment is chosen for the rearrangement process (Jung, Giallourakis et al. 2006). Transcription factor Pax5 is considered very important regarding commitment to mature B cell phenotype and regulating

the *IGHV* gene rearrangements. It regulates promoter contraction at *IGH* locus and mediates recombination of distal *IGHV* gene segments with rearranged *IGHDJ* gene segments. (Fuxa, Skok et al. 2004, Venigalla, McGuire et al. 2013).

Analysis of gene rearrangements taking place in the bone marrow during human B cell development indicates that non random usage of different *IGH* gene segments starts at an early stage. In early pre B cells when there is no surface Ig receptor, *IGHV*3-23, *IGHV*4-34 and *IGHV*4-59 dominate in the repertoire. In late pre B cells productively rearranged heavy chains pair with surrogate light chain and are expressed as pre BCR (Geier and Schlissel 2006, Meng, Yunk et al. 2011). It has been found that there are differences in the ability of different *IGH* gene segments to pair with surrogate light chain. Some IGH transcripts form stable complexes with surrogate light chain and preferentially proliferate. In this way additional biases in the usage of *IGH* gene segments are generated that are independent of the cellular ligands and antigens (Rao, Riggs et al. 1999, Meng, Yunk et al. 2011). However, it is not known if properties of the rearranged *IGH* influence subsequent binding to either IGK or IGL light chain and if this would affect the final IGH repertoire.

After successful *VDJ* gene rearrangement at the *IGH* locus, the light chain locus starts to rearrange *IGKV* and *IGKJ* gene segments first at the *IGK* locus and if this is not successful then at the *IGL* locus. After successful gene rearrangement of either light chain, IGH protein is expressed on cell surface with either IGK or IGL as a mature BCR. In order to maintain central tolerance and immunocompetency, BCRs are then exposed to ligands in the bone marrow that results in positive or negative selection of cells (von Muenchow, Engdahl et al. 2014). In this way certain *IGHV* gene segments are selected for or against in the repertoire. For example, in one study it was found that the *IGHV*3-11 gene segment was well represented at the pre B cell stage but not in the mature naïve B cells. On the other hand, *IGHV*3-20 gene segment was well represented in the mature naïve B cells but not in pre B cells (Kraj, Rao et al. 1997). This implies that certain *IGHV*, *IGKV* and *IGLV* genes are selected according to their specificities. This shows that biases in *IGH* gene segments usage can be generated at different stages of B cell development. However, it is not known if there is any role of light chain in the final expressed repertoire of IGH gene segments in BCRs

expressing either IGK or IGL light chains. Structurally, BCRs have both conserved FR and less-conserved CDRs. CDR-3 encompasses the junctions of rearranged IGH *VDJ* genes. As a consequence, the CDR-3 of IGH is highly variable in its physicochemical properties. Acidic and basic amino acids encoded by the CDR-3 sequence can affect the thermostability (Aliphatic index) and hydrophobicity (GRAVY index) of BCR. However, it is not known if the physicochemical properties of IGH peptide acquired during rearrangement can favour pairing of IGH with either IGK or IGL.

5.1.1 Aims of this chapter

It is not known if the properties of VDJ rearrangement at *IGH* locus influence the subsequent pairing with IGK or IGL. The aims of the investigation in this chapter were to ask:

1. Is there any preference for productively rearranged IGH having particular *IGHV*, *IGHD* or *IGHJ* gene segments to be expressed with IGK or IGL light chains?
2. Do physicochemical properties of CDR-3 region of IGH influence preferential pairing with IGK or IGL light chains?

5.2 Results

Lymphocytes from three anonymous blood donors were used for this study. PBMCs were isolated, washed and stored as viable cells in liquid nitrogen. Cells for sorting were thawed and stained using CD19-PerCP cy5.5, CD27-APC, IgD-PE-cy7, kappa-APC-cy7, lambda-pacific blue and live/dead-Am cyan and sorted into mature naïve B cells expressing IGK ($CD19^+IgD^+CD27^-IGK^+$) and mature naïve B cells expressing IGL ($CD19^+IgD^+CD27^-IGL^+$) as shown in Figure 5-1.

The number of cells derived from each donor that were used further for RNA isolation is indicated in Table 5-1.

Table 5-1: Number of mature naïve B cells isolated from individual donors.

Donor	Mature naïve B cells expressing IGK	Mature naïve B cells expressing IGL
HD-1	122,077	90,000
HD-2	101,342	66,409
HD-3	77,807	59,837

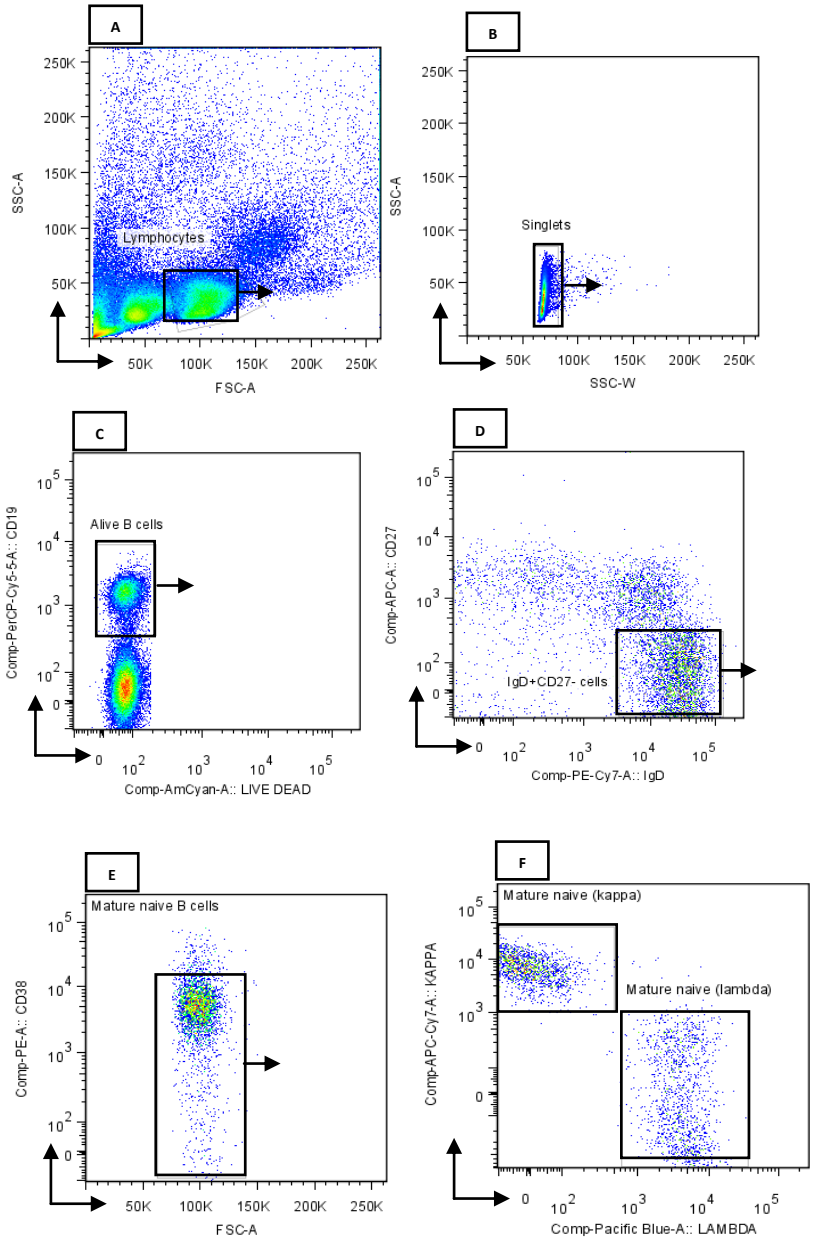


Figure 5-1: Example of sorting B cell subsets into mature naïve B cells expressing either IGK or IGL light chain. PBMCs were stained with monoclonal antibodies and sorted using FACS **(A)** Lymphocytes were gated first **(B)** Doublets were excluded **(C)** Gated on alive $CD19^{+}$ B cells **(D)** Gated on $IgD^{+}CD27^{-}$ B cells **(E)** Gated on mature naïve B cells by excluding $CD38^{hi}$ cells **(F)** Mature naïve B cells ($CD19^{+}CD27^{-}IgD^{+}CD38^{+/-}$) were sorted according to light chains (IGK or IGL)

Sorted cells were suspended in SLyRT buffer. Superscript enzyme was added and cDNA was generated by reverse transcription. After reverse transcription, cDNA was stored at -20ºC. In order to analyze the *IGH* repertoire of mature naïve B cells expressing IGK or IGL, *IGHVDJ* gene rearrangements were amplified from *IGVH* to I*GCμ* by semi nested PCR. MID tagged primers were used in the second round of PCR to distinguish amplicons generated from different individual donors and from the two subsets of sorted B cells expressing IGK or IGL. For this experiment six different MID tagged primers were used as follows:

1- Healthy donor 1
   - **a.** Mature naïve B cells expressing IGK (MID 3)
   - **b.** Mature naïve B cells expressing IGL (MID 4)

2- Healthy donor 2
   - **a.** Mature naïve B cells expressing IGK (MID 7)
   - **b.** Mature naïve B cells expressing IGL (MID 8)

3- Healthy donor 3
   - **a.** Mature naïve B cells expressing IGK (MID 11)
   - **b.** Mature naïve B cells expressing IGL (MID 12)

Raw data was received and our collaborators subjected the data to quality control as discussed in materials and methods (section 2.3.6.5). The number of unique IGH *VDJ* gene rearrangements from each donor is shown in Table 5-2. The number of sequences obtained from donor 2 was only 24 out of total 984 sequences derived from mature naïve B cells expressing IGK. This was a matter of concern since this could potentially bias the data when sequences from all individuals were pooled. This could cause artefacts since data is expressed as number of rearrangements of each family and the denominator for IGK and IGL expressing sequences was different for donor 2. However, pooling of sequences is common in studies such as this one and was therefore used with the caveat that anything that appeared to be important should also be apparent in each individual who contributed to the data.

Table 5-2: Number of unique DNA sequences derived from each donor. Numbers in parenthesis indicate the amount (nano grams) of cDNA used to generate these sequences after amplification.

Donor	IGH *VDJ* rearrangements from mature naïve B cells expressing IGK	IGH *VDJ* rearrangements from mature naïve B cells expressing IGL
HD-1	647 (450)	627 (370)
HD-2	24 (406)	384 (571)
HD-3	313 (351)	178 (389)
Total	984	1189

5.2.1 Is there any preference for productively rearranged IGH having particular *IGHV*, *IGHD* or *IGHJ* gene segments to be expressed with IGK or IGL light chains?

Variability in heavy chain peptide chain is encoded by *IGHV*, *IGHD* and *IGHJ* gene segments.

Usage of *IGHV* gene families

In order to determine if there is any preference for IGK or IGL to pair with certain *IGHV* gene families, *IGHV* family usage was compared between *IGH* genes of mature naïve B cells expressing either IGK or IGL. In pooled data from three individual donors, *IGHV*1 and *IGHV*2 were found to be associated significantly more with IGL. On the other hand *IGHV*3 and *IGHV6* paired significantly more with IGK. It was found that the most frequently used *IGHV* families in both groups were *IGHV*3 and *IGHV*4 from three pooled donors followed by *IGHV*1, *IGHV*2, *IGHV*5 and *IGHV*6 respectively (Figure 5-2).

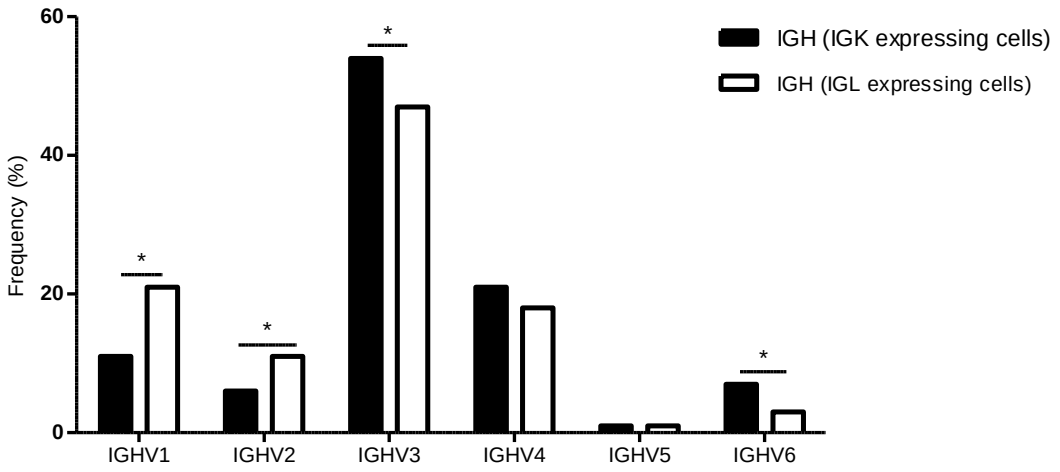


Figure 5-2: The frequencies of *IGHV* family usage. Comparison of relative rearrangement frequencies of *IGHV* gene families of in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values of ≤ 0.05 were considered significant (*). *IGHV1* and *IGHV2* families were used more frequently in IGL expressing B cells where as *IGHV3* and *IGHV6* families were used more commonly with IGK expressing B cells.

If the differences in *IGHV* family usage are generally biologically relevant, they should be observed in each donor individually. Therefore, data from each donor was analyzed separately (Figure 5-3). *IGHV1* family was found to be associated more with IGL in the two individuals that contributed most sequences, but this was not statistically significant in either case. There was a trend of more usage of *IGHV2* family with IGL in all three donors but it was not statistically significant. *IGHV3* family tended to be associated more with IGK in two individuals only. *IGHV6* family was expressed more with IGK in one person only however, usage was the same in both IGK and IGL expressing mature naïve B cells in the other two individuals. The significant features observed in the pooled data were therefore not all consistent between all individuals that contributed to the data. Since some significant biases were observed in the usage of *IGHV* families by IGK and IGL in pooled data but not in the

individual donors. Therefore, in order to better understand this, the relative usage of *IGHV* gene segments was analyzed.

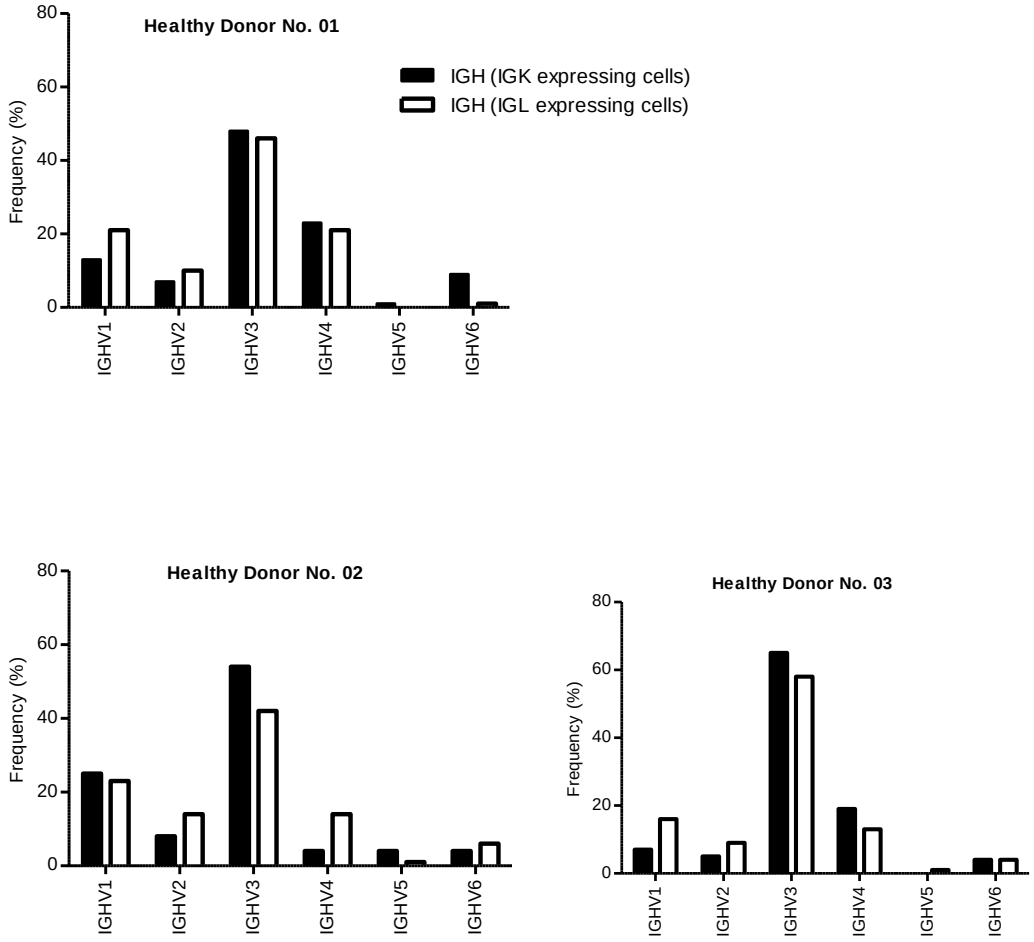


Figure 5-3: The relative frequency of *IGHV* family usage in three separate donors. Comparison of relative rearrangement frequencies of *IGHV* gene families of in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). Of the differences that appeared significant when all three datasets were pooled, only a tendency of *IGHV*2 to be more associated with IGL is apparent in all three individuals but without statistical significance.

Usage of *IGHV* gene segments

When usage of individual gene segments was analysed, some *IGHV* gene segments showed biases in preferential pairing with IGK or IGL. It was found that *IGHV*3-30-3 and *IGHV*6-1 gene segments were significantly associated more with the IGK. On the other hand, *IGHV*1-18 gene segment was preferentially paired with the IGL as shown in Figure 5-4.

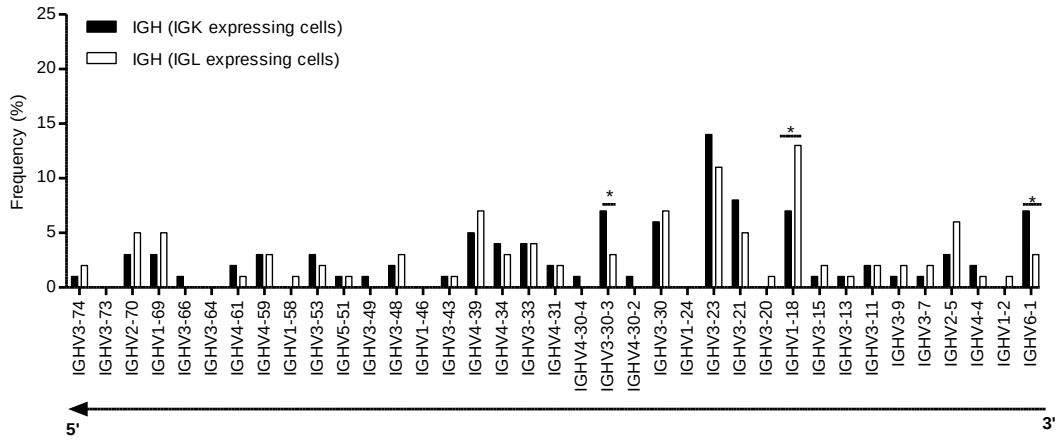


Figure 5-4: Analysis of *IGHV* segments usage. Comparison of relative rearrangement frequencies of *IGHV* gene segments in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). *IGHV*3-30-3 and *IGHV*6-1 segments were associated significantly more with IGK and *IGHV*1-18 was associated significantly more with IGL.

If the differences in *IGHV* family usage are consistent and biologically relevant and not a consequence of the sequence numbers that affect the denominators in the pooled data differently for IGK and IGL, they should be observed in each donor individually. Therefore, each individual was analyzed separately to look for biases in preferential pairing of *IGHV* segments with light chains (Figure 5-5). Significant biases in *IGHV* gene usage were observed in individuals, but these were confined only to that particular individual. Of the three significant differences that were observed in the pooled data, the *IGHV*6-1 gene segment was preferentially associated with IGK only in one individual while other two individuals had similar relative usage of *IGHV*6-1 in IGK and IGL expressing mature naïve B

cells. On the other hand, *IGHV*1-18 was found to be associated more with IGL in the two individuals that contributed most sequences to the overall data pool. This reached statistical significance for donor 1, and may therefore be a genuine finding. In contrast, the bias observed in the *IGHV*3-30 segment was only observed in the pooled data and not in individual donors and was an artefact of the data pooling with different sequence numbers.

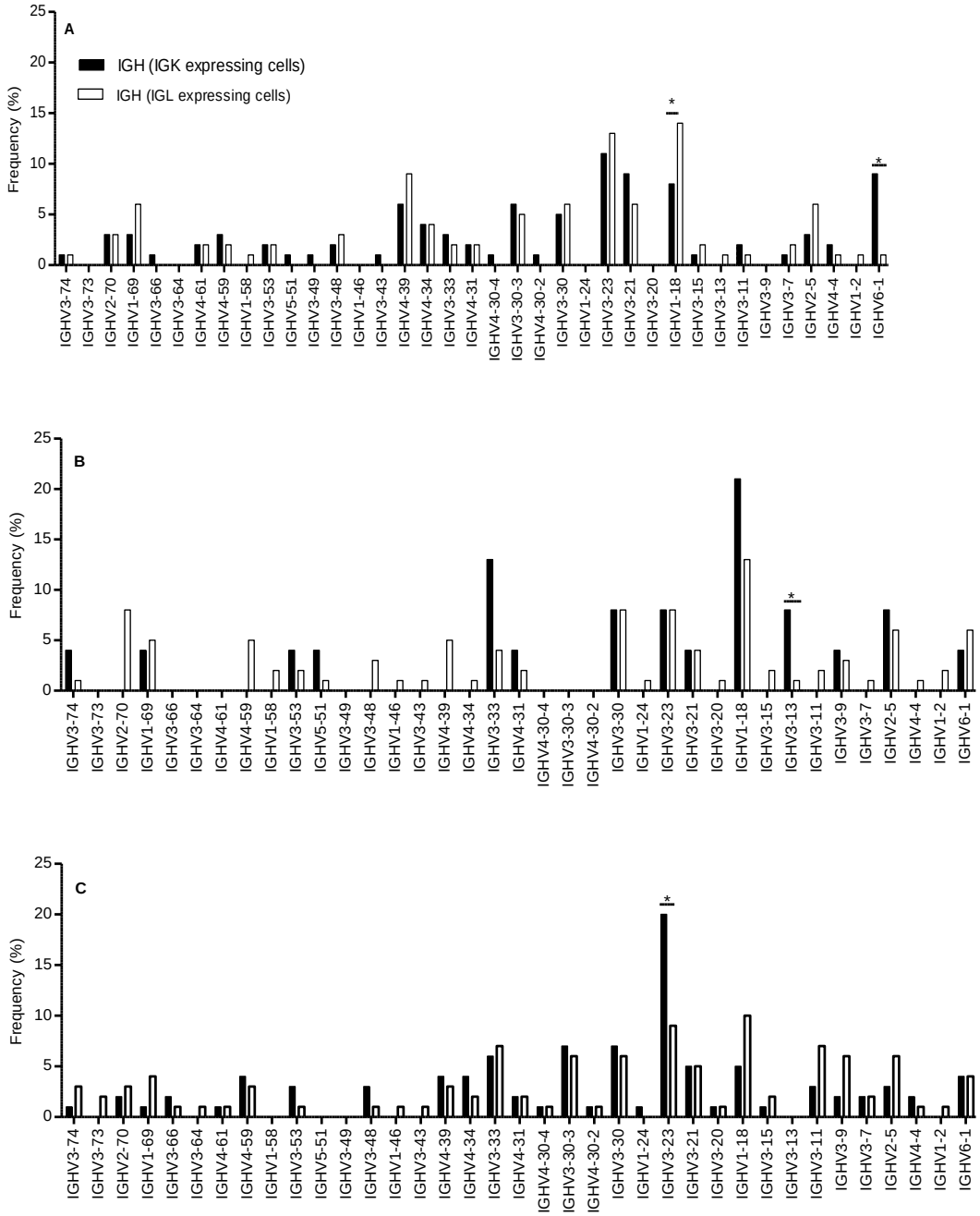


Figure 5-5: Analysis of *IGHV* segments usage in separate individuals (A-C). Comparison of relative rearrangement frequencies of *IGHV* gene segments in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). There were no consistent differences in the usage of *IGHV* segments by light chains.

Usage of *IGHJ* gene segments

There are 6 *IGHJ* gene segments. Relative usage of *IGHJ* genes in the expressed repertoire of mature naïve B cells expressing either IGK or IGL was compared. There was no difference in the usage of *IGHJ* gene families expressed with either IGK or IGL (Figure 5-6).

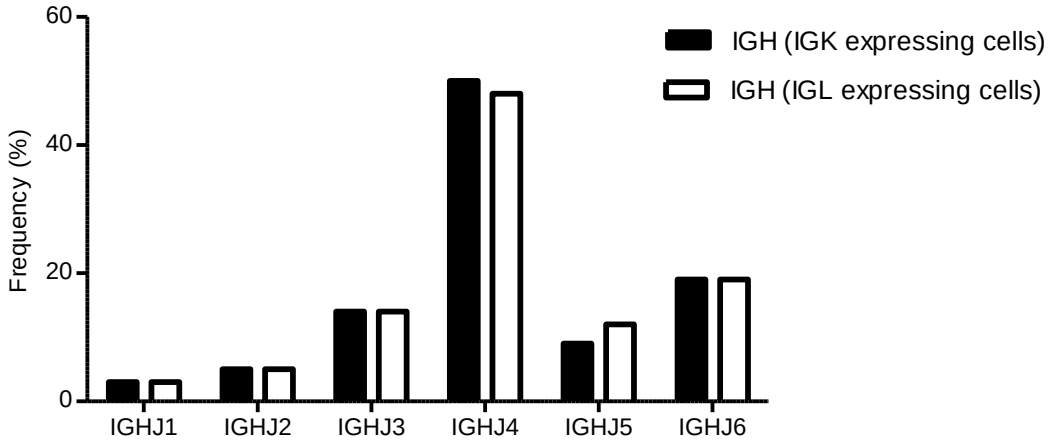


Figure 5-6: The frequencies of *IGHJ* family usage. Comparison of relative rearrangement frequencies of *IGHJ* gene families of in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). No differences were observed in the relative usage of *IGHJ* segments by IGK and IGL expressing mature naïve B cells.

Usage of *IGHD* gene families

There are 7 *IGHD* gene families to encode the variable region of heavy chain. Usage of *IGHD* families in association with light chains was compared. No differences were observed in the usage of *IGHD* families of heavy chain associated with either IGK or IGL (Figure 5-7).

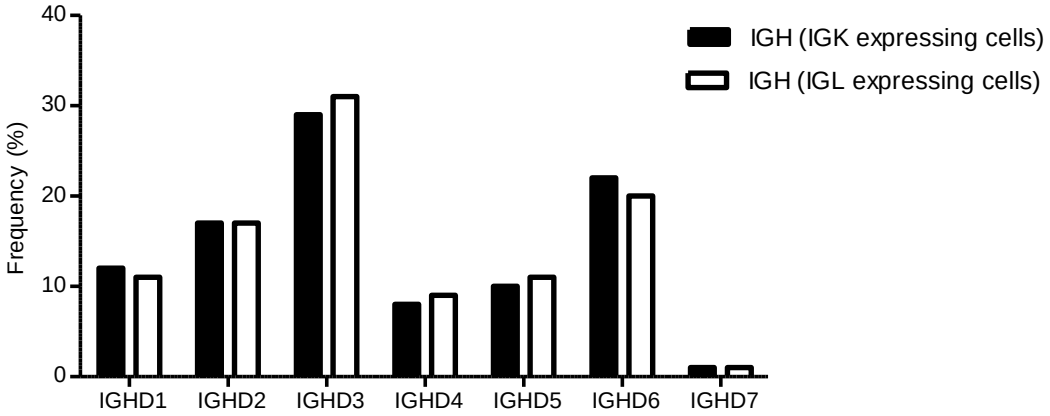


Figure 5-7: The frequencies of *IGHD* family usage. Comparison of relative rearrangement frequencies of *IGHD* gene families of in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). No differences were observed in the relative usage of *IGHD* gene families by IGK and IGL expressing mature naïve B cells.

Usage of *IGHD* gene segments

There are 23 *IGHD* gene segments that encode for the variable region of heavy chain. As noticed above there was no preference of IGL to be expressed with any particular *IGHD* family. In order to get more insight, usage of *IGHD* gene segments used in *IGH* VDJ rearrangements was compared from IGK or IGL expressing B cells. Frequency of usage of different *IGHD* genes in heavy chains associated with either IGK or IGL was also found to be similar (Figure 5-8).

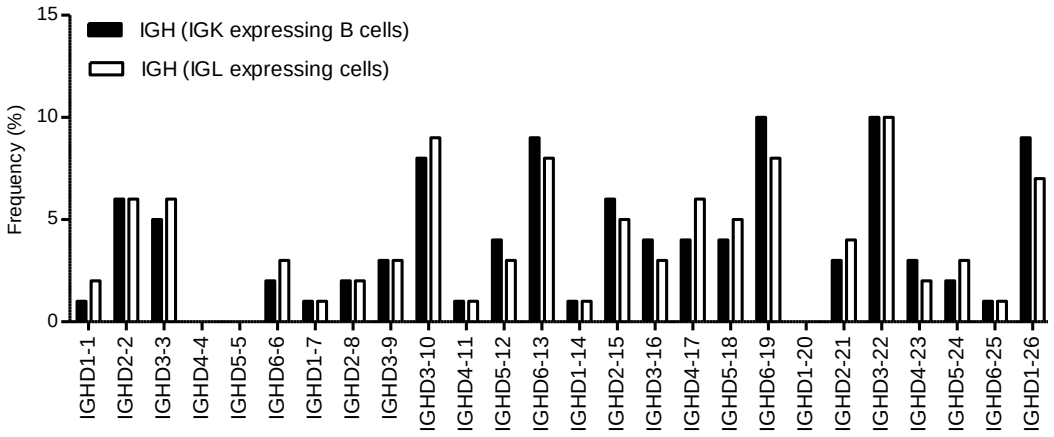


Figure 5-8: The frequencies of *IGHD* gene segments usage. Comparison of relative rearrangement frequencies of *IGHD* gene segments of in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*). No differences were observed in the relative usage of *IGHD* gene segments by IGK and IGL expressing mature naïve B cells.

5.2.2 Do physicochemical properties of CDR-3 of IGH influence preferential pairing with IGK or IGL light chains?

The gene rearrangement process generates highly diverse sequences in IGH and it was considered possible that certain features may favour the association with either IGK or IGL during B cell development. Therefore physicochemical properties of the heavy chain CDR-3 that contain the junctional region of rearranging segments and non templated inserted nucleotides were analysed in B cells expressing IGK or IGL.

CDR-3 length (Amino acids)

The CDR-3 length in the heavy chain genes of mature naïve B cells associated with either IGK or IGL ranged from 8 to 25 amino acids. The average length was 15 amino acids (Figure 5-9).

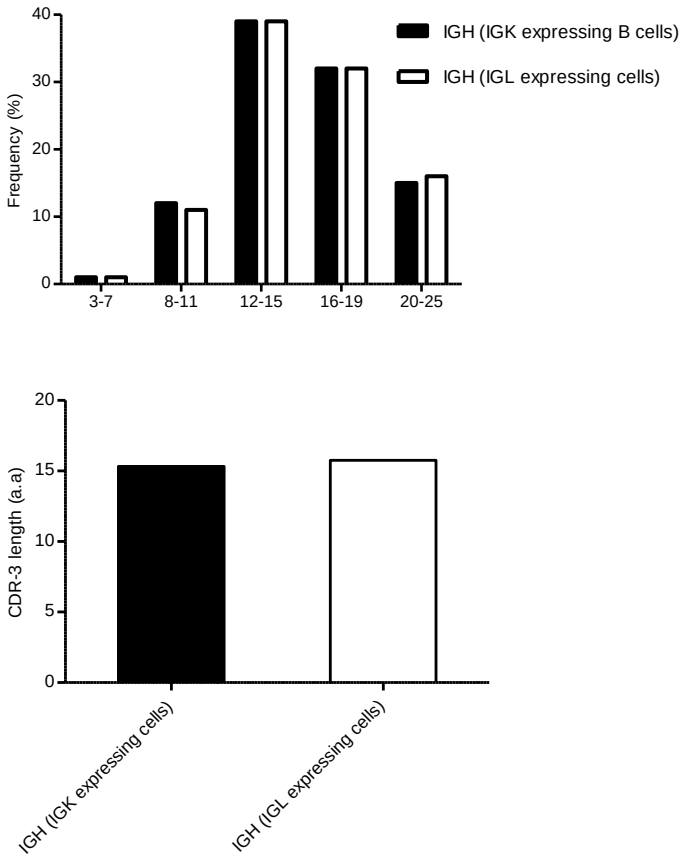


Figure 5-9: Average length of CDR-3 region in different B cell subsets. Non parametric tests were used to compare the means. No differences were observed in the CDR-3 length of IGH chain associated with either IGK or IGL chain.

N nucleotides insertion

During the rearrangement process at the IGH locus, random nucleotides can be inserted into the VH-DH and DH-JH junctions mediated by TdT. The N nucleotides at both junctions were pooled for the analysis. There was no difference in the number of nucleotides inserted in the IGH associated with either IGK or IGL. N nucleotides inserted at the junctions ranged from 5 to 30 nucleotides. The average number of nucleotides inserted at the junctions was 12 nucleotides (Figure 5-10).

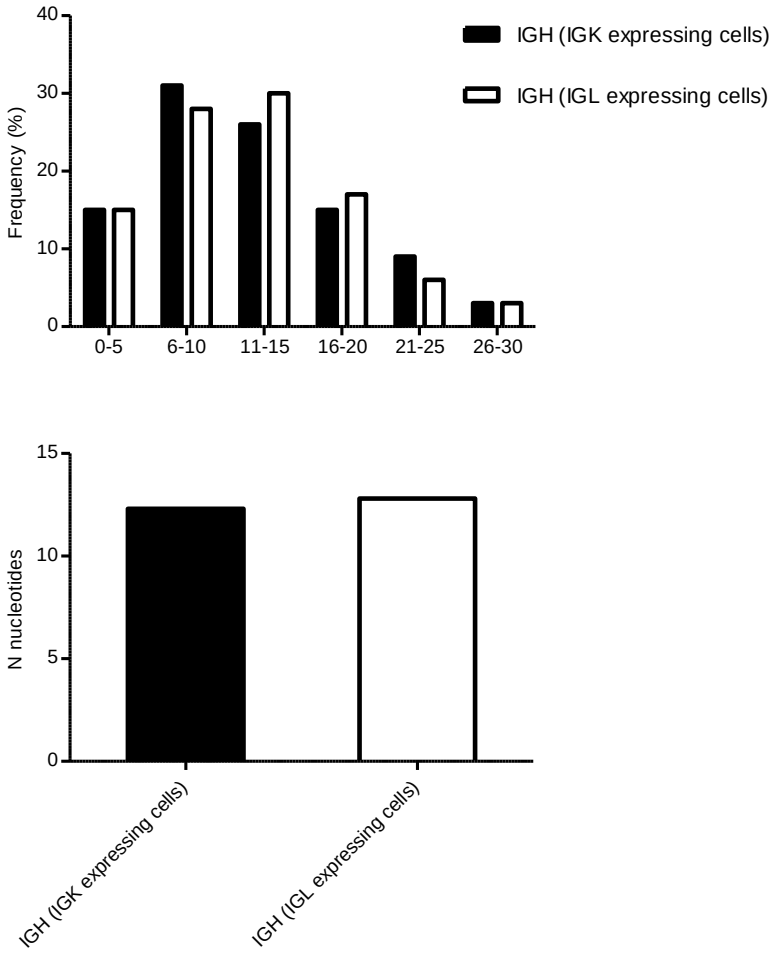


Figure 5-10: N nucleotides insertion in different B cell subsets. Non parametric tests were used to compare the means. No differences were observed in the N-nucleotides inserted in the IGH chain associated with either IGK or IGL chain.

Aliphatic and GRAVY indices

Aliphatic index is a measure of the relative volume occupied by aliphatic side chains (valine, isoleucine, and leucine) that has been regarded as a positive factor for the increase of thermostability of globular proteins. GRAVY index indicates the solubility of proteins (Kyte and Doolittle 1982). There was no difference in these indices of CDR-3 regions of IGH expressed with either IGK or IGL. The mean GRAVY index was -0.48 in mature naïve B cells expressing either IGK or IGL. The mean aliphatic index of CDR-3 of IGH associated with IGK or IGL was 56 (Figure 5-11).

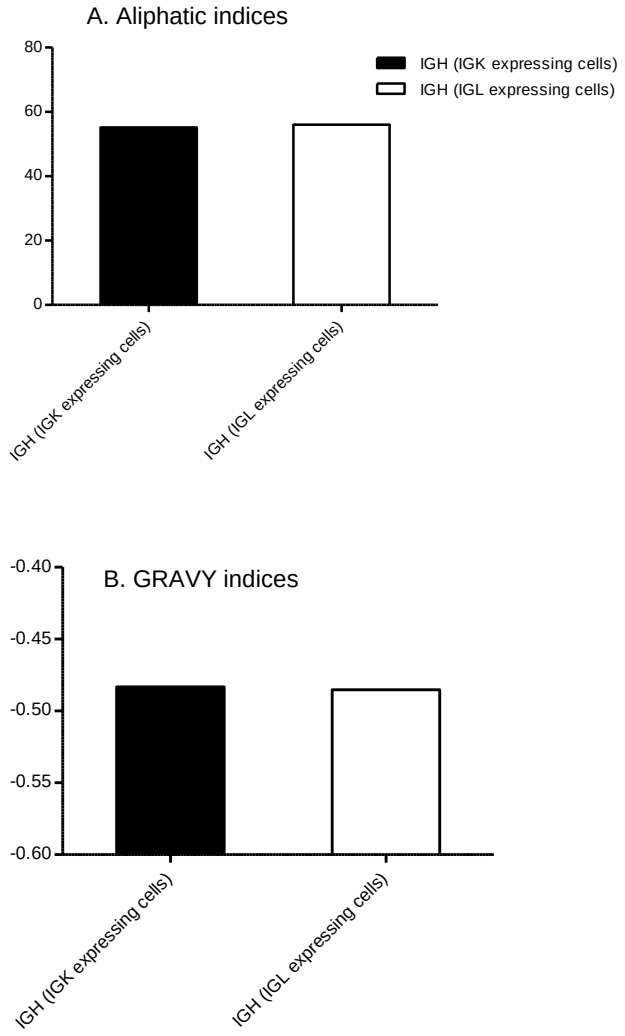


Figure 5-11: The aliphatic index (a) and GRAVY index (b) of CDR-3 regions of different B cell subsets. ProtPram was used to determine the aliphatic index of the CDR-3 region and the grand average hydropathy (GRAVY) index. Non parametric tests were used to compare the means. No differences were observed in the aliphatic index and GRAVY index of the IGH chain associated with either IGK or IGL chain.

Amino acids at the junction

Amino acid composition of the CDR-3 region was similar in mature naïve B subsets expressing either light chain. Negatively charged amino acids constituted up to 13% of the CDR-3 region while 10% were positively charged amino acids (Figure 5-12).

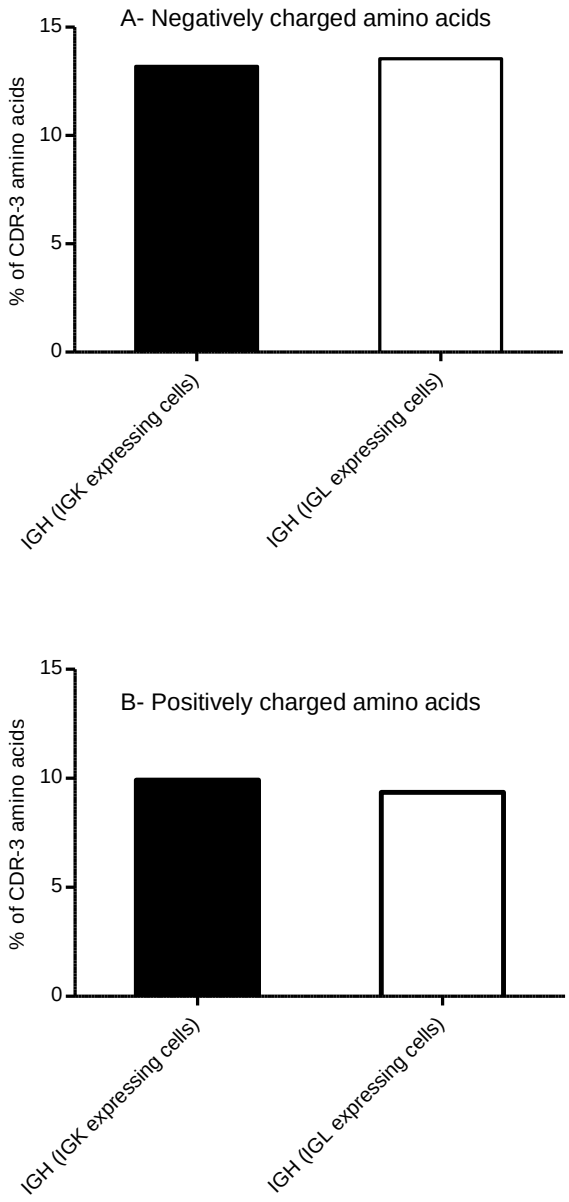


Figure 5-12: Amino acids composition of CDR-3 regions. Percentage of positive and negatively charged amino acids was calculated in the total number of amino acids in CDR-3. Non parametric test was used to compare the means. No differences were observed in the amino acid composition of the IGH chain associated with either IGK or IGL chain.

5.3 Discussion

The aim of this chapter was to ask whether properties of the *VDJ* rearrangement of the *IGH* bias subsequent choice of pairing with either IGK or IGL. To answer this question, 984 *IGH VDJ* gene rearrangements expressed with IGK were compared with 1189 IGH VDJ gene rearrangements expressed with IGL in the mature naïve B cells isolated from three healthy individuals. Essentially no consistent robust bias in pairing of *IGH* VDJ rearrangements with either IGK or IGL was observed.

Analysis of the pooled data from three individual donors revealed some differences in the usage of *IGHV* gene families in the *IGH* repertoire expressed with either IGK or IGL. *IGHV*3 and *IGHV*6 families were found to be expressed more with IGK on the other hand *IGHV*1 and *IGHV*2 were preferentially expressed more with IGL. In order to understand this better, usage of different *IGHV* gene segments was compared. It was found that *IGHV*3-30-3 and *IGHV*6-1 gene segments were expressed more with IGK while *IGHV*1-18 gene segment preferentially paired with IGL. However, most of these findings were not consistent when individuals were analyzed separately. The one exception was considered to be a possible bias towards pairing of *IGHV*1-18 in B cells using IGL. Evidence for such a bias was therefore sought in data from other groups. A study from Wardemann et al., (Wardemann, Yurasov et al. 2003) was able to identify heavy and light chain pairings in individual B cells by analysis of heavy and light chain sequences from single cells (Figure 5-13). When supplementary data from this paper was analyzed a slight bias towards the association of *IGHV*1-18 with IGL was observed. However, *IGHV*1-18 was not a commonly used segment in this dataset. Overall this indicates that preferential pairing of *IGHV* gene segments with light chains appears not to be a significant feature of human B cell development.

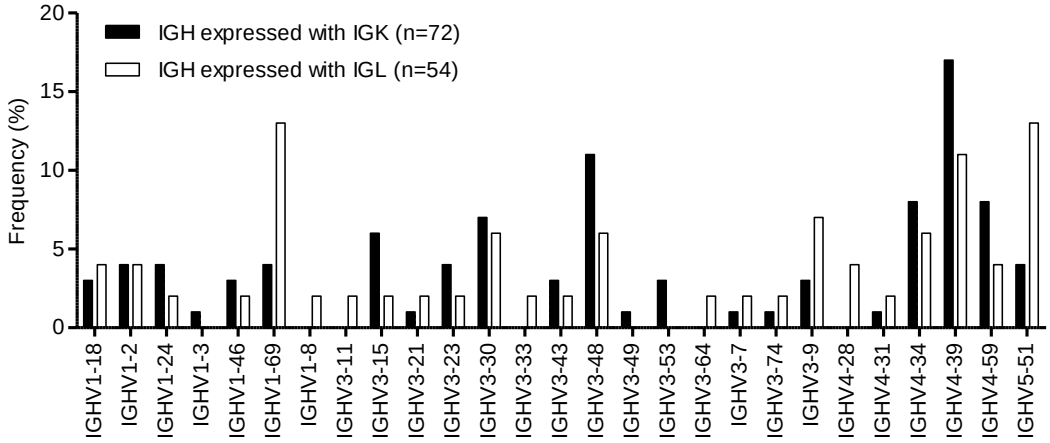


Figure 5-13: Analysis of supplementary data from Wardemann et al., (Wardemann, Yurasov et al. 2003) of single cells showing relative *IGH* gene usage by *IGK* and *IGL* gene segments. Comparison of relative rearrangement frequencies of *IGHV* gene segments in the expressed repertoire of mature naïve B cells associated with either IGK or IGL. Chi squared test was performed to compare the gene frequencies with Bonferroni post hoc test and p values with ≤ 0.05 were considered significant (*).

It is interesting that individual donors had significant tendencies to associate *IGHV* gene segments with IGK and IGL such as gene segment *IGHV*6-1 was significantly associated with IGK in one individual but in other two individuals there was no difference. Similarly *IGHV*3-23 gene segment was significantly associated more with IGK in one of the individuals. *IGHV*3-13 was also significantly associated with IGK in one individual while this segment was negligible in the expressed repertoire of other two individuals. Gene segment *IGHV*1-18 was found to be expressed more with IGL in two individuals. These differences in expression might reflect the selection of specific gene rearrangements in the expressed repertoire as this experiment was done using cDNAs rather than bias in the rearrangement process. There was a limitation in the data set that for one of the individuals (HD-2) number of sequences were not balanced between IGK and IGL as described in results section.

IGHD gene usage was also compared between IGH expressed with either IGK or IGL. The most frequently used family was *IGHD*3 followed by *IGHD*6 and *IGHD*2. However, there was no difference in the usage of *IGHD* family or individual gene segments in the IGH expressed with either IGK or IGL. In order to explore further and find any subtle differences usage of

IGHD gene segments was compared. The most commonly used *IGHD* segments were found to be *IGHD*3-10, *IGH*6-13, *IGHD*6-19, *IGHD*3-22 and *IGHD*1-26. However, *IGHD* gene segments were found to be used at the similar frequencies in IGH expressed with any light chain. *IGHJ* usage was compared between IGH expressed with different light chains. The most commonly used *IGHJ* family was *IGHJ*4 followed by *IGHJ*6 and *IGHJ*3. But there was no difference in the frequencies of usage of IGH expressed with either IGK or IGL. This implies that there is no influence of *IGH* VDJ repertoire in the selection of light chain during the course of B cell development.

The non templated junctional N and P nucleotides make a major contribution to the physicochemical properties of the CDR-3. It was therefore considered possible that certain physicochemical features could favour association with either IGK or IGL. However, there was no difference in the junctional properties such as CDR-3 length, amino acid composition, N- nucleotides insertion, and hydrophobicity and thermostability indices.

The literature shows that in some malignancies, antigens can select certain combinations of rearranged genes that can result in some biases in the expressed repertoire. In one study on CLL patients it was found that preferential pairing is determined by conserved CDR-3 region of *IGH* such as *IGHD*2-2/*IGHJ*6 encodes CLL69B motif that preferentially binds to the *IGKV*1-39. In other cases *IGHD*2-2/*IGHJ*6 encodes motif CLL69D that binds with *IGKV*3-1 and *IGHD*3-3/*IGHJ*6 encodes CLL69C that preferentially binds *IGL*3-9. In another study it was found *IGHV*3-21 is dominant and preferentially rearranged with *IGL*2-14 and *IGLJ*3. The finding of combined use of one certain *IGH* and light chain gene indicates that these segments have been preferentially rearranged in this cohort of CLL patients due to antigen mediated selection (Stamatopoulos, Belessi et al. 2005, Widhopf, Goldberg et al. 2008). However, this is not a feature of normal healthy B cells.

After the IGH peptide passes the checkpoint at pre BCR stage then gene recombination starts at the *IGK* locus. If the rearrangement is not successful at both *IGK* alleles then it results in the deactivation of *IGK* locus by KDE which allows the deletion of non functional *IGK* rearrangements. It is followed by gene rearrangement at the *IGL* locus.

5.4 Conclusions

Data in this chapter indicates that only non functionality at the *IGK* locus leads to gene rearrangement at the *IGL* locus. Following are the concluding points of this chapter:

1. There is no preference for productively rearranged IGH having particular *IGHV*, *IGHD* or *IGHJ* gene segments to be expressed with either IGK or IGL light chains. However, some biases may exist on individual basis.
2. The physicochemical properties of CDR-3 region of IGH do not influence preferential pairing with either IGK or IGL light chains.

Chapter 6

Overview and potential future directions

Conclusions

Overall in this thesis, some features of light chain gene rearrangements have been highlighted in the context of gene selection during B cell development. The following are the salient features:

1. Data in this thesis has identified that a cell can carry multiple gene rearrangements of light chain that can appear functional according to IMGT that may not be used. Solely investigation of gene rearrangements at the DNA level is not sufficient to provide an evidence of positive and negative selection of gene segments into the selected repertoire. Within each cell multiple DNA gene rearrangements can be retained. Some of these DNA gene rearrangements would potentially be expressed to constitute the BCR, however, if DNA gene rearrangements had been selected against during development then there could also have been the accumulation of these unused DNA gene rearrangements within a cell. However, such potentially expressed and unused DNA gene rearrangements cannot be differentiated after amplification by PCR because IMGT would display a productive CDR-3 junction for both and that can be misleading in the context of selection in the expressed repertoire. In chapter 4, several such aberrant productive gene rearrangements have been highlighted. Therefore, determination of relative frequencies of gene rearrangement at both DNA and RNA level can give better understanding of selection pressures of individual gene segments.

2. Gene rearrangements generated by editing may exist in the cell as episomal segments. When a cell divides, the episome will transfer to one of the daughter cells resulting in apparent biases in the repertoire. DNA editing is the prominent mechanism during establishment of naïve B cell repertoire. Thus, as a consequence of this process, several out-of-frame DNA gene rearrangements could accumulate inside the cell after excision from the chromosome. Previous studies have identified some biases in the ratios of in-frame and out-of frame DNA gene rearrangements at the *IGL* locus and attributed these to the characteristics of the

locus (Farner, Dorner et al. 1999, Richl, Stern et al. 2008). However, based on the data of chapter 3 of this thesis it can be suggested that the ratios of in-frame to out-of-frame gene rearrangements varies at different development stages of B cells. This was evident as there was relative abundance of out-of-frame DNA gene rearrangements in the naïve B cells as compared to IgA expressing B cells. It can be suggested there was a dilution of out-of frame excised circles as a result of multiplication during B cell maturation process. Thus, differences in the ratios of in-frame to out-of-frame gene rearrangements are more likely the reflection of developmental stage following cell division rather than a characteristic feature of the locus or gene segments.

High throughput sequencing experiments constituted an integral part of each chapter in this thesis. In order to retrieve a large number of sequences it was important to harvest a large number of cells from each sample. To meet this requirement, the strategy of using healthy buffy coat samples was employed. However, these buffy coats were procured from anonymous samples; therefore, the data was not available to us regarding ethnicity, clinical history, age, and gender. As these factors can modulate the immune system, utilization of these buffy coats imposed some limitations on the interpretation of the data.

Future directions

Chapter 3: It was observed that there was a significant reduction in the out-of-frame gene rearrangements in IgA expressing B cells. In future it would be interesting to know if the mucosal B cell subset in blood expressing higher levels of α4β7 integrin would have generated a different profile of gene rearrangements at the *IGL* locus. Additionally, changes in the surface expression of IGK:IGL in IgA subsets could be traced further to investigate if the reduction in IGK expressing cells observed in $CD27^{-}$IgA is a consequence of initiation of the secondary gene rearrangements at the *IGL* locus in IGK expressing $CD27^{-}$IgA cells.

Chapter 4: It was found that the profile of productive DNA gene rearrangements can be different from the expressed DNA gene rearrangements at the *IGK* locus. Moreover, productive gene rearrangement at the *IGK* locus is not the only factor for initiating DNA gene

rearrangements at the *IGL* locus and this feature has implications in the selection for or against of individual *IGKV* gene segments. For future studies selection processes could be compared with autoimmune diseases to determine if there is any defect in the selection of *IGKV* gene segments in pathological conditions. This study determined the *IGK* repertoire of naïve B cells; observations can also be extended by comparing naïve repertoire of *IGK* gene rearrangements with the repertoire of antigen experienced B cells in the periphery.

Chapter 5: There was no difference in the *IGH* gene rearrangements expressed in either IGK or IGL expressing naïve B cells. In future, it might be worthwhile to extend the investigation by comparing the expressed DNA gene rearrangements with genomic DNA gene rearrangements to determine positive and negative selection of individual gene segments.

References

- Acheson, D. W. and S. Luccioli (2004). "Microbial-gut interactions in health and disease. Mucosal immune responses." Best Pract Res Clin Gastroenterol **18**(2): 387-404.
- Ada, G. (2008). "The enunciation and impact of Macfarlane Burnet's clonal selection theory of acquired immunity." Immunol Cell Biol **86**(2): 116-118.
- Adams, E., A. Basten and C. C. Goodnow (1990). "Intrinsic B-cell hyporesponsiveness accounts for self-tolerance in lysozyme/anti-lysozyme double-transgenic mice." Proc Natl Acad Sci U S A **87**(15): 5687-5691.
- Aguilera, I., J. Melero, A. Nunez-Roldan and B. Sanchez (2001). "Molecular structure of eight human autoreactive monoclonal antibodies." Immunology **102**(3): 273-280.
- Akamatsu, Y., N. Tsurushita, F. Nagawa, M. Matsuoka, K. Okazaki, M. Imai and H. Sakano (1994). "Essential residues in V(D)J recombination signals." J Immunol **153**(10): 4520-4529.
- Allen, C. D., K. M. Ansel, C. Low, R. Lesley, H. Tamamura, N. Fujii and J. G. Cyster (2004). "Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5." Nat Immunol **5**(9): 943-952.
- Allen, C. D., T. Okada, H. L. Tang and J. G. Cyster (2007). "Imaging of germinal center selection events during affinity maturation." Science **315**(5811): 528-531.
- Alt, F. W., G. D. Yancopoulos, T. K. Blackwell, C. Wood, E. Thomas, M. Boss, R. Coffman, N. Rosenberg, S. Tonegawa and D. Baltimore (1984). "Ordered rearrangement of immunoglobulin heavy chain variable region segments." Embo j **3**(6): 1209-1219.
- Arbuckle, J. L., L. A. Fauss, R. Simpson, L. M. Ptaszek and K. K. Rodgers (2001). "Identification of two topologically independent domains in RAG1 and their role in macromolecular interactions relevant to V(D)J recombination." J Biol Chem **276**(40): 37093-37101.
- Atchison, M. L. (2014). "Function of YY1 in Long-Distance DNA Interactions." Front Immunol **5**: 45.

- Bankovich, A. J., S. Raunser, Z. S. Juo, T. Walz, M. M. Davis and K. C. Garcia (2007). "Structural insight into pre-B cell receptor function." Science **316**(5822): 291-294.
- Bannard, O., R. M. Horton, C. D. Allen, J. An, T. Nagasawa and J. G. Cyster (2013). "Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection." Immunity **39**(5): 912-924.
- Bargatze, R. F., M. A. Jutila and E. C. Butcher (1995). "Distinct roles of L-selectin and integrins alpha 4 beta 7 and LFA-1 in lymphocyte homing to Peyer's patch-HEV in situ: the multistep model confirmed and refined." Immunity **3**(1): 99-108.
- Barone, F., P. Patel, J. D. Sanderson and J. Spencer (2009). "Gut-associated lymphoid tissue contains the molecular machinery to support T-cell-dependent and T-cell-independent class switch recombination." Mucosal Immunol **2**(6): 495-503.
- Barone, F., A. Vossenkamper, L. Boursier, W. Su, A. Watson, S. John, D. K. Dunn-Walters, P. Fields, S. Wijetilleka, J. D. Edgeworth and J. Spencer (2011). "IgA-producing plasma cells originate from germinal centers that are induced by B-cell receptor engagement in humans." Gastroenterology **140**(3): 947-956.
- Bassing, C. H., W. Swat and F. W. Alt (2002). "The mechanism and regulation of chromosomal V(D)J recombination." Cell **109 Suppl**: S45-55.
- Beishuizen, A., M. A. Verhoeven, E. J. Mol and J. J. van Dongen (1994). "Detection of immunoglobulin kappa light-chain gene rearrangement patterns by Southern blot analysis." Leukemia **8**(12): 2228-2236; discussion 2237-2229.
- Benckert, J., N. Schmolka, C. Kreschel, M. J. Zoller, A. Sturm, B. Wiedenmann and H. Wardemann (2011). "The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific." J Clin Invest **121**(5): 1946-1955.
- Benichou, J., J. Glanville, E. T. Prak, R. Azran, T. C. Kuo, J. Pons, C. Desmarais, L. Tsaban and Y. Louzoun (2013). "The restricted DH gene reading frame usage in the expressed human antibody repertoire is selected based upon its amino acid content." J Immunol **190**(11): 5567-5577.

- Bergman, Y. and H. Cedar (2004). "A stepwise epigenetic process controls immunoglobulin allelic exclusion." Nat Rev Immunol **4**(10): 753-761.
- Bergqvist, P., E. Gardby, A. Stensson, M. Bemark and N. Y. Lycke (2006). "Gut IgA class switch recombination in the absence of CD40 does not occur in the lamina propria and is independent of germinal centers." J Immunol **177**(11): 7772-7783.
- Bergqvist, P., A. Stensson, N. Y. Lycke and M. Bemark (2010). "T cell-independent IgA class switch recombination is restricted to the GALT and occurs prior to manifest germinal center formation." J Immunol **184**(7): 3545-3553.
- Berkowska, M. A., G. J. Driessen, V. Bikos, C. Grosserichter-Wagener, K. Stamatopoulos, A. Cerutti, B. He, K. Biermann, J. F. Lange, M. van der Burg, J. J. van Dongen and M. C. van Zelm (2011). "Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways." Blood **118**(8): 2150-2158.
- Bernard, O., N. Hozumi and S. Tonegawa (1978). "Sequences of mouse immunoglobulin light chain genes before and after somatic changes." Cell **15**(4): 1133-1144.
- Bertocci, B., A. De Smet, J. C. Weill and C. A. Reynaud (2006). "Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo." Immunity **25**(1): 31-41.
- Bertrand, F. E., 3rd, L. G. Billips, P. D. Burrows, G. L. Gartland, H. Kubagawa and H. W. Schroeder, Jr. (1997). "Ig D(H) gene segment transcription and rearrangement before surface expression of the pan-B-cell marker CD19 in normal human bone marrow." Blood **90**(2): 736-744.
- Borsutzky, S., B. B. Cazac, J. Roes and C. A. Guzman (2004). "TGF-beta receptor signaling is critical for mucosal IgA responses." J Immunol **173**(5): 3305-3309.
- Bossen, C. and P. Schneider (2006). "BAFF, APRIL and their receptors: structure, function and signaling." Semin Immunol **18**(5): 263-275.
- Boyd, S. D., B. A. Gaeta, K. J. Jackson, A. Z. Fire, E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D.

Nguyen, K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder and A. M. Collins (2010). "Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements." J Immunol **184**(12): 6986-6992.

- Brady, B. L., N. C. Steinel and C. H. Bassing (2010). "Antigen receptor allelic exclusion: an update and reappraisal." J Immunol **185**(7): 3801-3808.
- Brandtzaeg, P. (2009). "Mucosal immunity: induction, dissemination, and effector functions." Scand J Immunol **70**(6): 505-515.
- Brandtzaeg, P. (2010). "Function of mucosa-associated lymphoid tissue in antibody formation." Immunol Invest **39**(4-5): 303-355.
- Brandtzaeg, P. (2013). "Secretory IgA: Designed for Anti-Microbial Defense." Front Immunol **4**: 222.
- Brandtzaeg, P. and F. E. Johansen (2005). "Mucosal B cells: phenotypic characteristics, transcriptional regulation, and homing properties." Immunol Rev **206**: 32-63.
- Brandtzaeg, P. and R. Pabst (2004). "Let's go mucosal: communication on slippery ground." Trends Immunol **25**(11): 570-577.
- Brauninger, A., T. Goossens, K. Rajewsky and R. Kuppers (2001). "Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human." Eur J Immunol **31**(12): 3631-3637.
- Bridges, S. L., Jr., S. K. Lee, M. L. Johnson, J. C. Lavelle, P. G. Fowler, W. J. Koopman and H. W. Schroeder, Jr. (1995). "Somatic mutation and CDR3 lengths of immunoglobulin kappa light chains expressed in patients with rheumatoid arthritis and in normal individuals." J Clin Invest **96**(2): 831-841.
- Briney, B. S., J. R. Willis, B. A. McKinney and J. E. Crowe, Jr. (2012). "High-throughput antibody sequencing reveals genetic evidence of global regulation of the naive and memory repertoires that extends across individuals." Genes Immun **13**(6): 469-473.
- Casola, S. and K. Rajewsky (2006). "B cell recruitment and selection in mouse GALT germinal centers." Curr Top Microbiol Immunol **308**: 155-171.

- Castigli, E., S. Scott, F. Dedeoglu, P. Bryce, H. Jabara, A. K. Bhan, E. Mizoguchi and R. S. Geha (2004). "Impaired IgA class switching in APRIL-deficient mice." Proc Natl Acad Sci U S A **101**(11): 3903-3908.
- Castigli, E., S. A. Wilson, L. Garibyan, R. Rachid, F. Bonilla, L. Schneider and R. S. Geha (2005). "TACI is mutant in common variable immunodeficiency and IgA deficiency." Nat Genet **37**(8): 829-834.
- Cerutti, A. (2008). "The regulation of IgA class switching." Nat Rev Immunol **8**(6): 421-434.
- Cerutti, A. and M. Rescigno (2008). "The biology of intestinal immunoglobulin A responses." Immunity **28**(6): 740-750.
- Chakalova, L., E. Debrand, J. A. Mitchell, C. S. Osborne and P. Fraser (2005). "Replication and transcription: shaping the landscape of the genome." Nat Rev Genet **6**(9): 669-677.
- Chaudhuri, J. and F. W. Alt (2004). "Class-switch recombination: interplay of transcription, DNA deamination and DNA repair." Nat Rev Immunol **4**(7): 541-552.
- Chen, C., Z. Nagy, M. Z. Radic, R. R. Hardy, D. Huszar, S. A. Camper and M. Weigert (1995). "The site and stage of anti-DNA B-cell deletion." Nature **373**(6511): 252-255.
- Chen, C., E. L. Prak and M. Weigert (1997). "Editing disease-associated autoantibodies." Immunity **6**(1): 97-105.
- Collins, A. M., Y. Wang, V. Singh, P. Yu, K. J. Jackson and W. A. Sewell (2008). "The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate." Immunogenetics **60**(11): 669-676.
- Corcoran, A. E. (2010). "The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation." Semin Immunol **22**(6): 353-361.
- Cortes, P., Z. S. Ye and D. Baltimore (1994). "RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1." Proc Natl Acad Sci U S A **91**(16): 7633-7637.

- Crago, S. S., W. H. Kutteh, I. Moro, M. R. Allansmith, J. Radl, J. J. Haaijman and J. Mestecky (1984). "Distribution of IgA1-, IgA2-, and J chain-containing cells in human tissues." J Immunol **132**(1): 16-18.
- Curry, J. D., J. K. Geier and M. S. Schlissel (2005). "Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis." Nat Immunol **6**(12): 1272-1279.
- Das, S., N. Nikolaidis and M. Nei (2009). "Genomic organization and evolution of immunoglobulin kappa gene enhancers and kappa deleting element in mammals." Mol Immunol **46**(15): 3171-3177.
- Davi, F., A. Faili, C. Gritti, C. Blanc, C. Laurent, L. Sutton, C. Schmitt and H. Merle-Beral (1997). "Early onset of immunoglobulin heavy chain gene rearrangements in normal human bone marrow CD34+ cells." Blood **90**(10): 4014-4021.
- De, P. and K. K. Rodgers (2004). "Putting the pieces together: identification and characterization of structural domains in the V(D)J recombination protein RAG1." Immunol Rev **200**: 70-82.
- De Silva, N. S. and U. Klein (2015). "Dynamics of B cells in germinal centres." Nat Rev Immunol.
- Desiderio, S. V., G. D. Yancopoulos, M. Paskind, E. Thomas, M. A. Boss, N. Landau, F. W. Alt and D. Baltimore (1984). "Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells." Nature **311**(5988): 752-755.
- Di Noia, J. M. and M. S. Neuberger (2007). "Molecular mechanisms of antibody somatic hypermutation." Annu Rev Biochem **76**: 1-22.
- Difilippantonio, M. J., C. J. McMahan, Q. M. Eastman, E. Spanopoulou and D. G. Schatz (1996). "RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination." Cell **87**(2): 253-262.
- Dorner, T., S. J. Foster, N. L. Farner and P. E. Lipsky (1998). "Immunoglobulin kappa chain receptor editing in systemic lupus erythematosus." J Clin Invest **102**(4): 688-694.

- Dreyer, W. J. and J. C. Bennett (1965). "The molecular basis of antibody formation: a paradox." Proc Natl Acad Sci U S A **54**(3): 864-869.
- Dullaers, M., D. Li, Y. Xue, L. Ni, I. Gayet, R. Morita, H. Ueno, K. A. Palucka, J. Banchereau and S. Oh (2009). "A T cell-dependent mechanism for the induction of human mucosal homing immunoglobulin A-secreting plasmablasts." Immunity **30**(1): 120-129.
- Dworzak, M. N., G. Fritsch, G. Froschl, D. Printz and H. Gadner (1998). "Four-color flow cytometric investigation of terminal deoxynucleotidyl transferase-positive lymphoid precursors in pediatric bone marrow: CD79a expression precedes CD19 in early B-cell ontogeny." Blood **92**(9): 3203-3209.
- Early, P., H. Huang, M. Davis, K. Calame and L. Hood (1980). "An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH." Cell **19**(4): 981-992.
- Edelman, G. M. and B. Benacerraf (1962). "On structural and functional relations between antibodies and proteins of the gamma-system." Proc Natl Acad Sci U S A **48**: 1035-1042.
- Ehlich, A., V. Martin, W. Muller and K. Rajewsky (1994). "Analysis of the B-cell progenitor compartment at the level of single cells." Curr Biol **4**(7): 573-583.
- Ehlich, A., S. Schaal, H. Gu, D. Kitamura, W. Muller and K. Rajewsky (1993). "Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development." Cell **72**(5): 695-704.
- Elson, C. O. and W. Ealding (1984). "Generalized systemic and mucosal immunity in mice after mucosal stimulation with cholera toxin." J Immunol **132**(6): 2736-2741.
- Fagarasan, S. and T. Honjo (2003). "Intestinal IgA synthesis: regulation of front-line body defences." Nat Rev Immunol **3**(1): 63-72.
- Fagarasan, S., K. Kinoshita, M. Muramatsu, K. Ikuta and T. Honjo (2001). "In situ class switching and differentiation to IgA-producing cells in the gut lamina propria." Nature **413**(6856): 639-643.

- Farner, N. L., T. Dorner and P. E. Lipsky (1999). "Molecular mechanisms and selection influence the generation of the human V lambda J lambda repertoire." J Immunol **162**(4): 2137-2145.
- Feddersen, R. M., D. J. Martin and B. G. Van Ness (1990). "The frequency of multiple recombination events occurring at the human Ig kappa L chain locus." J Immunol **144**(3): 1088-1093.
- Feddersen, R. M., D. J. Martin and B. G. Van Ness (1990). "Novel recombinations of the IG kappa-locus that result in allelic exclusion." J Immunol **145**(2): 745-750.
- Feddersen, R. M. and B. G. Van Ness (1985). "Double recombination of a single immunoglobulin kappa-chain allele: implications for the mechanism of rearrangement." Proc Natl Acad Sci U S A **82**(14): 4793-4797.
- Feeney, A. J., A. Tang and K. M. Ogwaro (2000). "B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization." Immunol Rev **175**: 59-69.
- Flanagan, J. G. and T. H. Rabbitts (1982). "Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes." Nature **300**(5894): 709-713.
- Foster, S. J., H. P. Brezinschek, R. I. Brezinschek and P. E. Lipsky (1997). "Molecular mechanisms and selective influences that shape the kappa gene repertoire of IgM+ B cells." J Clin Invest **99**(7): 1614-1627.
- Fugmann, S. D., A. I. Lee, P. E. Shockett, I. J. Villey and D. G. Schatz (2000). "The RAG proteins and V(D)J recombination: complexes, ends, and transposition." Annu Rev Immunol **18**: 495-527.
- Fulcher, D. A. and A. Basten (1994). "Reduced life span of anergic self-reactive B cells in a double-transgenic model." J Exp Med **179**(1): 125-134.
- Fuxa, M., J. Skok, A. Souabni, G. Salvagiotto, E. Roldan and M. Busslinger (2004). "Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene." Genes Dev **18**(4): 411-422.

- Galy, A., M. Travis, D. Cen and B. Chen (1995). "Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset." Immunity **3**(4): 459-473.
- Gatto, D., D. Paus, A. Basten, C. R. Mackay and R. Brink (2009). "Guidance of B cells by the orphan G protein-coupled receptor EBI2 shapes humoral immune responses." Immunity **31**(2): 259-269.
- Gauss, G. H. and M. R. Lieber (1992). "The basis for the mechanistic bias for deletional over inversional V(D)J recombination." Genes Dev **6**(8): 1553-1561.
- Gay, D., T. Saunders, S. Camper and M. Weigert (1993). "Receptor editing: an approach by autoreactive B cells to escape tolerance." J Exp Med **177**(4): 999-1008.
- Geier, J. K. and M. S. Schlissel (2006). "Pre-BCR signals and the control of Ig gene rearrangements." Semin Immunol **18**(1): 31-39.
- Geissmann, F., P. Launay, B. Pasquier, Y. Lepelletier, M. Leborgne, A. Lehuen, N. Brousse and R. C. Monteiro (2001). "A subset of human dendritic cells expresses IgA Fc receptor (CD89), which mediates internalization and activation upon cross-linking by IgA complexes." J Immunol **166**(1): 346-352.
- Gellert, M. (2002). "V(D)J recombination: RAG proteins, repair factors, and regulation." Annu Rev Biochem **71**: 101-132.
- Ghia, P., A. Gratwohl, E. Signer, T. H. Winkler, F. Melchers and A. G. Rolink (1995). "Immature B cells from human and mouse bone marrow can change their surface light chain expression." Eur J Immunol **25**(11): 3108-3114.
- Ghia, P., E. ten Boekel, E. Sanz, A. de la Hera, A. Rolink and F. Melchers (1996). "Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci." J Exp Med **184**(6): 2217-2229.
- Giachino, C., E. Padovan and A. Lanzavecchia (1995). "kappa+lambda+ dual receptor B cells are present in the human peripheral repertoire." J Exp Med **181**(3): 1245-1250.

- Giachino, C., E. Padovan and A. Lanzavecchia (1998). "Re-expression of RAG-1 and RAG-2 genes and evidence for secondary rearrangements in human germinal center B lymphocytes." Eur J Immunol **28**(11): 3506-3513.
- Gibbons, D. L. and J. Spencer (2011). "Mouse and human intestinal immunity: same ballpark, different players; different rules, same score." Mucosal Immunol **4**(2): 148-157.
- Girschick, H. J., A. C. Grammer, T. Nanki, M. Mayo and P. E. Lipsky (2001). "RAG1 and RAG2 expression by B cell subsets from human tonsil and peripheral blood." J Immunol **166**(1): 377-386.
- Gitlin, A. D., Z. Shulman and M. C. Nussenzweig (2014). "Clonal selection in the germinal centre by regulated proliferation and hypermutation." Nature **509**(7502): 637-640.
- Glanville, J., T. C. Kuo, H. C. von Budingen, L. Guey, J. Berka, P. D. Sundar, G. Huerta, G. R. Mehta, J. R. Oksenberg, S. L. Hauser, D. R. Cox, A. Rajpal and J. Pons (2011). "Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation." Proc Natl Acad Sci U S A **108**(50): 20066-20071.
- Goldmit, M., Y. Ji, J. Skok, E. Roldan, S. Jung, H. Cedar and Y. Bergman (2005). "Epigenetic ontogeny of the Igk locus during B cell development." Nat Immunol **6**(2): 198-203.
- Goodnow, C. C., J. Crosbie, S. Adelstein, T. B. Lavoie, S. J. Smith-Gill, R. A. Brink, H. Pritchard-Briscoe, J. S. Wotherspoon, R. H. Loblay, K. Raphael and et al. (1988). "Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice." Nature **334**(6184): 676-682.
- Goodnow, C. C., J. Crosbie, S. Adelstein, T. B. Lavoie, S. J. Smith-Gill, R. A. Brink, H. Pritchard-Briscoe, J. S. Wotherspoon, R. H. Loblay, K. Raphael, R. J. Trent and A. Basten (2009). "Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice." J Immunol **183**(9): 5442-5448.
- Grandien, A., R. Fucs, A. Nobrega, J. Andersson and A. Coutinho (1994). "Negative selection of multireactive B cell clones in normal adult mice." Eur J Immunol **24**(6): 1345-1352.

- Graninger, W. B., P. L. Goldman, C. C. Morton, S. J. O'Brien and S. J. Korsmeyer (1988). "The kappa-deleting element. Germline and rearranged, duplicated and dispersed forms." J Exp Med **167**(2): 488-501.
- Grimaldi, C. M., R. Hicks and B. Diamond (2005). "B cell selection and susceptibility to autoimmunity." J Immunol **174**(4): 1775-1781.
- Gugler, E., G. Bokelmann, A. Datwyler and G. Von Muralt (1958). "[Immunoelectrophoretic studies on human milk proteins]." Schweiz Med Wochenschr **88**(50): 1264-1267.
- Hall, J. A., J. R. Grainger, S. P. Spencer and Y. Belkaid (2011). "The role of retinoic acid in tolerance and immunity." Immunity **35**(1): 13-22.
- Halverson, R., R. M. Torres and R. Pelanda (2004). "Receptor editing is the main mechanism of B cell tolerance toward membrane antigens." Nat Immunol **5**(6): 645-650.
- Hamel, K. M., M. Mandal, S. Karki and M. R. Clark (2014). "Balancing Proliferation with Igkappa Recombination during B-lymphopoiesis." Front Immunol **5**: 139.
- Han, S., S. R. Dillon, B. Zheng, M. Shimoda, M. S. Schlissel and G. Kelsoe (1997). "V(D)J recombinase activity in a subset of germinal center B lymphocytes." Science **278**(5336): 301-305.
- Han, S., B. Zheng, D. G. Schatz, E. Spanopoulou and G. Kelsoe (1996). "Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells." Science **274**(5295): 2094-2097.
- Hanson, L. A. and P. Brandtzaeg (1993). "The discovery of secretory IgA and the mucosal immune system." Immunol Today **14**(8): 416-417.
- Hardenberg, G., L. Planelles, C. M. Schwarte, L. van Bostelen, T. Le Huong, M. Hahne and J. P. Medema (2007). "Specific TLR ligands regulate APRIL secretion by dendritic cells in a PKR-dependent manner." Eur J Immunol **37**(10): 2900-2911.
- Hardy, R. R. and K. Hayakawa (2001). "B cell development pathways." Annu Rev Immunol **19**: 595-621.
- Harrison, E. H. and M. M. Hussain (2001). "Mechanisms involved in the intestinal digestion and absorption of dietary vitamin A." J Nutr **131**(5): 1405-1408.

- Hart, A. L., S. C. Ng, E. Mann, H. O. Al-Hassi, D. Bernardo and S. C. Knight (2010). "Homing of immune cells: role in homeostasis and intestinal inflammation." Inflamm Bowel Dis **16**(11): 1969-1977.
- Hartley, S. B., M. P. Cooke, D. A. Fulcher, A. W. Harris, S. Cory, A. Basten and C. C. Goodnow (1993). "Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death." Cell **72**(3): 325-335.
- Hartley, S. B., J. Crosbie, R. Brink, A. B. Kantor, A. Basten and C. C. Goodnow (1991). "Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens." Nature **353**(6346): 765-769.
- Hayakawa, K., M. Asano, S. A. Shinton, M. Gui, L. J. Wen, J. Dashoff and R. R. Hardy (2003). "Positive selection of anti-thy-1 autoreactive B-1 cells and natural serum autoantibody production independent from bone marrow B cell development." J Exp Med **197**(1): 87-99.
- He, B., W. Xu, P. A. Santini, A. D. Polydorides, A. Chiu, J. Estrella, M. Shan, A. Chadburn, V. Villanacci, A. Plebani, D. M. Knowles, M. Rescigno and A. Cerutti (2007). "Intestinal bacteria trigger T cell-independent immunoglobulin A(2) class switching by inducing epithelial-cell secretion of the cytokine APRIL." Immunity **26**(6): 812-826.
- Heremans, J. F., M. T. Heremans and H. E. Schultze (1959). "Isolation and description of a few properties of the beta 2A-globulin of human serum." Clin Chim Acta **4**(1): 96-102.
- Hertz, M., V. Kouskoff, T. Nakamura and D. Nemazee (1998). "V(D)J recombinase induction in splenic B lymphocytes is inhibited by antigen-receptor signalling." Nature **394**(6690): 292-295.
- Hesse, J. E., M. R. Lieber, K. Mizuuchi and M. Gellert (1989). "V(D)J recombination: a functional definition of the joining signals." Genes Dev **3**(7): 1053-1061.
- Hieter, P. A., S. J. Korsmeyer, T. A. Waldmann and P. Leder (1981). "Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells." Nature **290**(5805): 368-372.

- Hikida, M. and H. Ohmori (1998). "Rearrangement of lambda light chain genes in mature B cells in vitro and in vivo. Function of reexpressed recombination-activating gene (RAG) products." J Exp Med **187**(5): 795-799.
- Hiom, K. and M. Gellert (1998). "Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination." Mol Cell **1**(7): 1011-1019.
- Hombach, J., F. Lottspeich and M. Reth (1990). "Identification of the genes encoding the IgM-alpha and Ig-beta components of the IgM antigen receptor complex by amino-terminal sequencing." Eur J Immunol **20**(12): 2795-2799.
- Hooper, L. V., D. R. Littman and A. J. Macpherson (2012). "Interactions between the microbiota and the immune system." Science **336**(6086): 1268-1273.
- Hozumi, N. and S. Tonegawa (1976). "Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions." Proc Natl Acad Sci U S A **73**(10): 3628-3632.
- Hsu, L. Y., H. E. Liang, K. Johnson, C. Kang and M. S. Schlissel (2004). "Pax5 activates immunoglobulin heavy chain V to DJ rearrangement in transgenic thymocytes." J Exp Med **199**(6): 825-830.
- Islam, K. B., L. Nilsson, P. Sideras, L. Hammarstrom and C. I. Smith (1991). "TGF-beta 1 induces germ-line transcripts of both IgA subclasses in human B lymphocytes." Int Immunol **3**(11): 1099-1106.
- Iwata, M., A. Hirakiyama, Y. Eshima, H. Kagechika, C. Kato and S. Y. Song (2004). "Retinoic acid imprints gut-homing specificity on T cells." Immunity **21**(4): 527-538.
- Jackson, K. J., M. J. Kidd, Y. Wang and A. M. Collins (2013). "The shape of the lymphocyte receptor repertoire: lessons from the B cell receptor." Front Immunol **4**: 263.
- Jiang, H., F. C. Chang, A. E. Ross, J. Lee, K. Nakayama, K. Nakayama and S. Desiderio (2005). "Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle." Mol Cell **18**(6): 699-709.
- Johnson, K., D. L. Pflugh, D. Yu, D. G. Hesslein, K. I. Lin, A. L. Bothwell, A. Thomas-Tikhonenko, D. G. Schatz and K. Calame (2004). "B cell-specific loss of

histone 3 lysine 9 methylation in the V(H) locus depends on Pax5." Nat Immunol **5**(8): 853-861.

- Johnston, C. M., A. L. Wood, D. J. Bolland and A. E. Corcoran (2006). "Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region." J Immunol **176**(7): 4221-4234.
- Jones, J. M. and M. Gellert (2002). "Ordered assembly of the V(D)J synaptic complex ensures accurate recombination." Embo j **21**(15): 4162-4171.
- Jung, D., C. Giallourakis, R. Mostoslavsky and F. W. Alt (2006). "Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus." Annu Rev Immunol **24**: 541-570.
- Kadaoui, K. A. and B. Corthesy (2007). "Secretory IgA mediates bacterial translocation to dendritic cells in mouse Peyer's patches with restriction to mucosal compartment." J Immunol **179**(11): 7751-7757.
- Kato, L. M., S. Kawamoto, M. Maruya and S. Fagarasan (2014). "Gut TFH and IgA: key players for regulation of bacterial communities and immune homeostasis." Immunol Cell Biol **92**(1): 49-56.
- Kawasaki, K., S. Minoshima, E. Nakato, K. Shibuya, A. Shintani, J. L. Schmeits, J. Wang and N. Shimizu (1997). "One-megabase sequence analysis of the human immunoglobulin lambda gene locus." Genome Research **7**(3): 250-261.
- Keenan, R. A., A. De Riva, B. Corleis, L. Hepburn, S. Licence, T. H. Winkler and I. L. Martensson (2008). "Censoring of autoreactive B cell development by the pre-B cell receptor." Science **321**(5889): 696-699.
- Kerfoot, S. M., G. Yaari, J. R. Patel, K. L. Johnson, D. G. Gonzalez, S. H. Kleinstein and A. M. Haberman (2011). "Germinal center B cell and T follicular helper cell development initiates in the interfollicular zone." Immunity **34**(6): 947-960.
- Kett, K., P. Brandtzaeg, J. Radl and J. J. Haaijman (1986). "Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues." J Immunol **136**(10): 3631-3635.

- Keyna, U., G. B. Beck-Engeser, J. Jongstra, S. E. Applequist and H. M. Jack (1995). "Surrogate light chain-dependent selection of Ig heavy chain V regions." J Immunol **155**(12): 5536-5542.
- Kilian, M., J. Reinholdt, H. Lomholt, K. Poulsen and E. V. Frandsen (1996). "Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence." Apmis **104**(5): 321-338.
- Kim, M. S., M. Lapkouski, W. Yang and M. Gellert (2015). "Crystal structure of the V(D)J recombinase RAG1-RAG2." Nature **518**(7540): 507-511.
- Kinoshita, K., M. Harigai, S. Fagarasan, M. Muramatsu and T. Honjo (2001). "A hallmark of active class switch recombination: transcripts directed by I promoters on looped-out circular DNAs." Proc Natl Acad Sci U S A **98**(22): 12620-12623.
- Kitano, M., S. Moriyama, Y. Ando, M. Hikida, Y. Mori, T. Kurosaki and T. Okada (2011). "Bcl6 protein expression shapes pre-germinal center B cell dynamics and follicular helper T cell heterogeneity." Immunity **34**(6): 961-972.
- Kline, G. H., L. Hartwell, G. B. Beck-Engeser, U. Keyna, S. Zaharevitz, N. R. Klinman and H. M. Jack (1998). "Pre-B cell receptor-mediated selection of pre-B cells synthesizing functional mu heavy chains." J Immunol **161**(4): 1608-1618.
- Klobeck, H. G. and H. G. Zachau (1986). "The human CK gene segment and the kappa deleting element are closely linked." Nucleic Acids Res **14**(11): 4591-4603.
- Komori, T., A. Okada, V. Stewart and F. W. Alt (1993). "Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes." Science **261**(5125): 1171-1175.
- Korsmeyer, S. J., P. A. Hieter, S. O. Sharrow, C. K. Goldman, P. Leder and T. A. Waldmann (1982). "Normal human B cells display ordered light chain gene rearrangements and deletions." J Exp Med **156**(4): 975-985.
- Kosak, S. T., J. A. Skok, K. L. Medina, R. Riblet, M. M. Le Beau, A. G. Fisher and H. Singh (2002). "Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development." Science **296**(5565): 158-162.

- Kotnis, A., L. Du, C. Liu, S. W. Popov and Q. Pan-Hammarstrom (2009). "Non-homologous end joining in class switch recombination: the beginning of the end." Philos Trans R Soc Lond B Biol Sci **364**(1517): 653-665.
- Kraj, P., S. P. Rao, A. M. Glas, R. R. Hardy, E. C. Milner and L. E. Silberstein (1997). "The human heavy chain Ig V region gene repertoire is biased at all stages of B cell ontogeny, including early pre-B cells." J Immunol **158**(12): 5824-5832.
- Krangel, M. S. (2009). "Mechanics of T cell receptor gene rearrangement." Curr Opin Immunol **21**(2): 133-139.
- Kumar, R., L. J. DiMenna, J. Chaudhuri and T. Evans (2014). "Biological function of activation-induced Cytidine Deaminase (AID)." Biomed J.
- Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-132.
- Lam, K. P., R. Kuhn and K. Rajewsky (1997). "In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death." Cell **90**(6): 1073-1083.
- Langerak, A. W., B. Nadel, A. De Torbal, I. L. Wolvers-Tettero, E. J. van Gastel-Mol, B. Verhaaf, U. Jager and J. J. van Dongen (2004). "Unraveling the consecutive recombination events in the human IGK locus." J Immunol **173**(6): 3878-3888.
- Larimore, K., M. W. McCormick, H. S. Robins and P. D. Greenberg (2012). "Shaping of human germline IgH repertoires revealed by deep sequencing." J Immunol **189**(6): 3221-3230.
- Lautner-Rieske, A., C. Huber, A. Meindl, W. Pargent, K. F. Schable, R. Thiebe, I. Zocher and H. G. Zachau (1992). "The human immunoglobulin kappa locus. Characterization of the duplicated A regions." Eur J Immunol **22**(4): 1023-1029.
- LeBien, T. W. (2000). "Fates of human B-cell precursors." Blood **96**(1): 9-23.
- LeBien, T. W., B. Wormann, J. G. Villablanca, C. L. Law, L. M. Steinberg, V. O. Shah and M. R. Loken (1990). "Multiparameter flow cytometric analysis of human fetal bone marrow B cells." Leukemia **4**(5): 354-358.
- Lee, J. and S. Desiderio (1999). "Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair." Immunity **11**(6): 771-781.

- Lefranc, M. P., V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc and S. Kossida (2015). "IMGT(R), the international ImMunoGeneTics information system(R) 25 years on." Nucleic Acids Res **43**(Database issue): D413-422.
- Levesque, M. C., M. A. Moody, K. K. Hwang, D. J. Marshall, J. F. Whitesides, J. D. Amos, T. C. Gurley, S. Allgood, B. B. Haynes, N. A. Vandergrift, S. Plonk, D. C. Parker, M. S. Cohen, G. D. Tomaras, P. A. Goepfert, G. M. Shaw, J. E. Schmitz, J. J. Eron, N. J. Shaheen, C. B. Hicks, H. X. Liao, M. Markowitz, G. Kelsoe, D. M. Margolis and B. F. Haynes (2009). "Polyclonal B cell differentiation and loss of gastrointestinal tract germinal centers in the earliest stages of HIV-1 infection." PLoS Med **6**(7): e1000107.
- Lewis, M. J., R. J. Pleass, M. R. Batten, J. D. Atkin and J. M. Woof (2005). "Structural requirements for the interaction of human IgA with the human polymeric Ig receptor." J Immunol **175**(10): 6694-6701.
- Lin, M., L. Du, P. Brandtzaeg and Q. Pan-Hammarstrom (2014). "IgA subclass switch recombination in human mucosal and systemic immune compartments." Mucosal Immunol **7**(3): 511-520.
- Lin, W. C. and S. Desiderio (1994). "Cell cycle regulation of V(D)J recombination-activating protein RAG-2." Proc Natl Acad Sci U S A **91**(7): 2733-2737.
- Litinskiy, M. B., B. Nardelli, D. M. Hilbert, B. He, A. Schaffer, P. Casali and A. Cerutti (2002). "DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL." Nat Immunol **3**(9): 822-829.
- Liu, Y., R. Subrahmanyam, T. Chakraborty, R. Sen and S. Desiderio (2007). "A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement." Immunity **27**(4): 561-571.
- Loken, M. R., V. O. Shah, K. L. Dattilio and C. I. Civin (1987). "Flow cytometric analysis of human bone marrow. II. Normal B lymphocyte development." Blood **70**(5): 1316-1324.

- Luning Prak, E. T., M. Monestier and R. A. Eisenberg (2011). "B cell receptor editing in tolerance and autoimmunity." Ann N Y Acad Sci **1217**: 96-121.
- Lutz, J., M. R. Heideman, E. Roth, P. van den Berk, W. Muller, C. Raman, M. Wabl, H. Jacobs and H. M. Jack (2011). "Pro-B cells sense productive immunoglobulin heavy chain rearrangement irrespective of polypeptide production." Proc Natl Acad Sci U S A **108**(26): 10644-10649.
- Lutz, J., W. Muller and H. M. Jack (2006). "VH replacement rescues progenitor B cells with two nonproductive VDJ alleles." J Immunol **177**(10): 7007-7014.
- Ma, Y., U. Pannicke, K. Schwarz and M. R. Lieber (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." Cell **108**(6): 781-794.
- Mackay, F., P. Schneider, P. Rennert and J. Browning (2003). "BAFF AND APRIL: a tutorial on B cell survival." Annu Rev Immunol **21**: 231-264.
- Macpherson, A. J. (2006). "IgA adaptation to the presence of commensal bacteria in the intestine." Curr Top Microbiol Immunol **308**: 117-136.
- Macpherson, A. J. and N. L. Harris (2004). "Interactions between commensal intestinal bacteria and the immune system." Nat Rev Immunol **4**(6): 478-485.
- Macpherson, A. J., L. Hunziker, K. McCoy and A. Lamarre (2001). "IgA responses in the intestinal mucosa against pathogenic and non-pathogenic microorganisms." Microbes Infect **3**(12): 1021-1035.
- Macpherson, A. J., K. D. McCoy, F. E. Johansen and P. Brandtzaeg (2008). "The immune geography of IgA induction and function." Mucosal Immunol **1**(1): 11-22.
- Macpherson, A. J. and E. Slack (2007). "The functional interactions of commensal bacteria with intestinal secretory IgA." Curr Opin Gastroenterol **23**(6): 673-678.
- Mandal, M., S. E. Powers, M. Maienschein-Cline, E. T. Bartom, K. M. Hamel, B. L. Kee, A. R. Dinner and M. R. Clark (2011). "Epigenetic repression of the Igk locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2." Nat Immunol **12**(12): 1212-1220.

- Mandik-Nayak, L., A. Bui, H. Noorchashm, A. Eaton and J. Erikson (1997). "Regulation of anti-double-stranded DNA B cells in nonautoimmune mice: localization to the T-B interface of the splenic follicle." J Exp Med **186**(8): 1257-1267.
- Martin, D., R. Q. Huang, T. LeBien and B. Van Ness (1991). "Induced rearrangement of kappa genes in the BLIN-1 human pre-B cell line correlates with germline J-C kappa and V kappa transcription." J Exp Med **173**(3): 639-645.
- Matthews, A. G., A. J. Kuo, S. Ramon-Maiques, S. Han, K. S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D. N. Ciccone, K. L. Walter, P. J. Utz, Y. Shi, T. G. Kutateladze, W. Yang, O. Gozani and M. A. Oettinger (2007). "RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination." Nature **450**(7172): 1106-1110.
- McBlane, J. F., D. C. van Gent, D. A. Ramsden, C. Romeo, C. A. Cuomo, M. Gellert and M. A. Oettinger (1995). "Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps." Cell **83**(3): 387-395.
- McMahan, C. J., M. J. Difilippantonio, N. Rao, E. Spanopoulou and D. G. Schatz (1997). "A basic motif in the N-terminal region of RAG1 enhances V(D)J recombination activity." Mol Cell Biol **17**(8): 4544-4552.
- McMurry, M. T. and M. S. Krangel (2000). "A role for histone acetylation in the developmental regulation of VDJ recombination." Science **287**(5452): 495-498.
- Meffre, E., E. Davis, C. Schiff, C. Cunningham-Rundles, L. B. Ivashkiv, L. M. Staudt, J. W. Young and M. C. Nussenzweig (2000). "Circulating human B cells that express surrogate light chains and edited receptors." Nat Immunol **1**(3): 207-213.
- Meffre, E., F. Papavasiliou, P. Cohen, O. de Bouteiller, D. Bell, H. Karasuyama, C. Schiff, J. Banchereau, Y. J. Liu and M. C. Nussenzweig (1998). "Antigen receptor engagement turns off the V(D)J recombination machinery in human tonsil B cells." J Exp Med **188**(4): 765-772.
- Melchers, F., H. Karasuyama, D. Haasner, S. Bauer, A. Kudo, N. Sakaguchi, B. Jameson and A. Rolink (1993). "The surrogate light chain in B-cell development." Immunol Today **14**(2): 60-68.

- Melchers, F., E. ten Boekel, T. Seidl, X. C. Kong, T. Yamagami, K. Onishi, T. Shimizu, A. G. Rolink and J. Andersson (2000). "Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells." Immunol Rev **175**: 33-46.
- Meng, W., L. Yunk, L. S. Wang, A. Maganty, E. Xue, P. L. Cohen, R. A. Eisenberg, M. G. Weigert, S. J. Mancini and E. T. Prak (2011). "Selection of individual VH genes occurs at the pro-B to pre-B cell transition." J Immunol **187**(4): 1835-1844.
- Meru, N., A. Jung, I. Baumann and G. Niedobitek (2002). "Expression of the recombination-activating genes in extrafollicular lymphocytes but no apparent reinduction in germinal center reactions in human tonsils." Blood **99**(2): 531-537.
- Mic, F. A., A. Molotkov, D. M. Benbrook and G. Duester (2003). "Retinoid activation of retinoic acid receptor but not retinoid X receptor is sufficient to rescue lethal defect in retinoic acid synthesis." Proc Natl Acad Sci U S A **100**(12): 7135-7140.
- Miqueu, P., M. Guillet, N. Degauque, J. C. Dore, J. P. Soulillou and S. Brouard (2007). "Statistical analysis of CDR3 length distributions for the assessment of T and B cell repertoire biases." Mol Immunol **44**(6): 1057-1064.
- Mo, X., T. Bailin and M. J. Sadofsky (2001). "A C-terminal region of RAG1 contacts the coding DNA during V(D)J recombination." Mol Cell Biol **21**(6): 2038-2047.
- Monestier, M. and M. Zouali (2002). "Receptor revision and systemic lupus." Scand J Immunol **55**(5): 425-431.
- Montalbano, A., K. M. Ogwaro, A. Tang, A. G. Matthews, M. Larijani, M. A. Oettinger and A. J. Feeney (2003). "V(D)J recombination frequencies can be profoundly affected by changes in the spacer sequence." J Immunol **171**(10): 5296-5304.
- Mouquet, H., J. F. Scheid, M. J. Zoller, M. Krogsgaard, R. G. Ott, S. Shukair, M. N. Artyomov, J. Pietzsch, M. Connors, F. Pereyra, B. D. Walker, D. D. Ho, P. C. Wilson, M. S. Seaman, H. N. Eisen, A. K. Chakraborty, T. J. Hope, J. V. Ravetch, H. Wardemann and M. C. Nussenzweig (2010). "Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation." Nature **467**(7315): 591-595.
- Mundy, C. L., N. Patenge, A. G. Matthews and M. A. Oettinger (2002). "Assembly of the RAG1/RAG2 synaptic complex." Mol Cell Biol **22**(1): 69-77.

- Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai and T. Honjo (2000). "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme." Cell **102**(5): 553-563.
- Muramatsu, M., H. Nagaoka, R. Shinkura, N. A. Begum and T. Honjo (2007). "Discovery of activation-induced cytidine deaminase, the engraver of antibody memory." Adv Immunol **94**: 1-36.
- Murre, C. (2005). "Helix-loop-helix proteins and lymphocyte development." Nat Immunol **6**(11): 1079-1086.
- Nadel, B., A. Tang, G. Escuro, G. Lugo and A. J. Feeney (1998). "Sequence of the spacer in the recombination signal sequence affects V(D)J rearrangement frequency and correlates with nonrandom Vkappa usage in vivo." J Exp Med **187**(9): 1495-1503.
- Nagafuchi, H., H. Yoshikawa, Y. Takeba, K. Nara, K. Miura, M. S. Kurokawa and N. Suzuki (2004). "Recombination activating genes (RAG) induce secondary Ig gene rearrangement in and subsequent apoptosis of human peripheral blood circulating B lymphocytes." Clin Exp Immunol **136**(1): 76-84.
- Nagawa, F., S. Hirose, H. Nishizumi, T. Nishihara and H. Sakano (2004). "Joining mutants of RAG1 and RAG2 that demonstrate impaired interactions with the coding-end DNA." J Biol Chem **279**(37): 38360-38368.
- Nayar, S., L. Paklet, J. Spencer and W. Su (2010). "Inactivation of unused alleles of human immunoglobulin light chain genes as a mechanism of self-preservation." Mol Immunol **47**(6): 1171-1172.
- Nemazee, D. (2006). "Receptor editing in lymphocyte development and central tolerance." Nat Rev Immunol **6**(10): 728-740.
- Nemazee, D. and K. Buerki (1989). "Clonal deletion of autoreactive B lymphocytes in bone marrow chimeras." Proc Natl Acad Sci U S A **86**(20): 8039-8043.
- Nemazee, D. and M. Weigert (2000). "Revising B cell receptors." J Exp Med **191**(11): 1813-1817.

- Neuberger, M. S., H. M. Caskey, S. Pettersson, G. T. Williams and M. A. Surani (1989). "Isotype exclusion and transgene down-regulation in immunoglobulin-lambda transgenic mice." Nature **338**(6213): 350-352.
- Norvell, A., L. Mandik and J. G. Monroe (1995). "Engagement of the antigen-receptor on immature murine B lymphocytes results in death by apoptosis." J Immunol **154**(9): 4404-4413.
- Nossal, G. J. and B. L. Pike (1980). "Clonal anergy: persistence in tolerant mice of antigen-binding B lymphocytes incapable of responding to antigen or mitogen." Proc Natl Acad Sci U S A **77**(3): 1602-1606.
- Nossal, G. J. and B. L. Pike (2007). "Evidence for the clonal abortion theory of B-lymphocyte tolerance. 1975." J Immunol **179**(9): 5619-5632.
- Oettinger, M. A. (2004). "How to keep V(D)J recombination under control." Immunol Rev **200**: 165-181.
- Oettinger, M. A., D. G. Schatz, C. Gorka and D. Baltimore (1990). "RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination." Science **248**(4962): 1517-1523.
- Okada, T., M. J. Miller, I. Parker, M. F. Krummel, M. Neighbors, S. B. Hartley, A. O'Garra, M. D. Cahalan and J. G. Cyster (2005). "Antigen-engaged B cells undergo chemotaxis toward the T zone and form motile conjugates with helper T cells." PLoS Biol **3**(6): e150.
- Okamoto, M., M. Murakami, A. Shimizu, S. Ozaki, T. Tsubata, S. Kumagai and T. Honjo (1992). "A transgenic model of autoimmune hemolytic anemia." J Exp Med **175**(1): 71-79.
- Owen, R. L. (1977). "Sequential uptake of horseradish peroxidase by lymphoid follicle epithelium of Peyer's patches in the normal unobstructed mouse intestine: an ultrastructural study." Gastroenterology **72**(3): 440-451.
- Palm, N. W., M. R. de Zoete, T. W. Cullen, N. A. Barry, J. Stefanowski, L. Hao, P. H. Degnan, J. Hu, I. Peter, W. Zhang, E. Ruggiero, J. H. Cho, A. L. Goodman and R. A. Flavell (2014). "Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease." Cell **158**(5): 1000-1010.

- Pan, X., M. Papasani, Y. Hao, M. Calamito, F. Wei, W. J. Quinn Iii, A. Basu, J. Wang, S. Hodawadekar, K. Zaprazna, H. Liu, Y. Shi, D. Allman, M. Cancro and M. L. Atchison (2013). "YY1 controls Igkappa repertoire and B-cell development, and localizes with condensin on the Igkappa locus." EMBO J **32**(8): 1168-1182.
- Pargent, W., A. Meindl, R. Thiebe, S. Mitzel and H. G. Zachau (1991). "The human immunoglobulin kappa locus. Characterization of the duplicated O regions." Eur J Immunol **21**(8): 1821-1827.
- Pasquier, B., P. Launay, Y. Kanamaru, I. C. Moura, S. Pfirsch, C. Ruffie, D. Henin, M. Benhamou, M. Pretolani, U. Blank and R. C. Monteiro (2005). "Identification of FcalphaRI as an inhibitory receptor that controls inflammation: dual role of FcRgamma ITAM." Immunity **22**(1): 31-42.
- Paus, D., T. G. Phan, T. D. Chan, S. Gardam, A. Basten and R. Brink (2006). "Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation." J Exp Med **203**(4): 1081-1091.
- Pauza, M. E., J. A. Rehmann and T. W. LeBien (1993). "Unusual patterns of immunoglobulin gene rearrangement and expression during human B cell ontogeny: human B cells can simultaneously express cell surface kappa and lambda light chains." J Exp Med **178**(1): 139-149.
- Pelanda, R. (2014). "Dual immunoglobulin light chain B cells: Trojan horses of autoimmunity?" Curr Opin Immunol **27**: 53-59.
- Pelanda, R., S. Schwers, E. Sonoda, R. M. Torres, D. Nemazee and K. Rajewsky (1997). "Receptor editing in a transgenic mouse model: site, efficiency, and role in B cell tolerance and antibody diversification." Immunity **7**(6): 765-775.
- Pelanda, R. and R. M. Torres (2012). "Central B-cell tolerance: where selection begins." Cold Spring Harb Perspect Biol **4**(4): a007146.
- Pernis, B., G. Chiappino, A. S. Kelus and P. G. Gell (1965). "Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues." J Exp Med **122**(5): 853-876.

- Peterson, D. A., N. P. McNulty, J. L. Guruge and J. I. Gordon (2007). "IgA response to symbiotic bacteria as a mediator of gut homeostasis." Cell Host Microbe **2**(5): 328-339.
- Phalipon, A., A. Cardona, J. P. Kraehenbuhl, L. Edelman, P. J. Sansonetti and B. Corthesy (2002). "Secretory component: a new role in secretory IgA-mediated immune exclusion in vivo." Immunity **17**(1): 107-115.
- Pleass, R. J., M. L. Lang, M. A. Kerr and J. M. Woof (2007). "IgA is a more potent inducer of NADPH oxidase activation and degranulation in blood eosinophils than IgE." Mol Immunol **44**(6): 1401-1408.
- Prak, E. L., M. Trounstine, D. Huszar and M. Weigert (1994). "Light chain editing in kappa-deficient animals: a potential mechanism of B cell tolerance." J Exp Med **180**(5): 1805-1815.
- Puget, N., R. Hirasawa, N. S. Hu, N. Laviolette-Malirat, R. Feil and A. A. Khamlichi (2015). "Insertion of an imprinted insulator into the IgH locus reveals developmentally regulated, transcription-dependent control of V(D)J recombination." Mol Cell Biol **35**(3): 529-543.
- Putnam, F. W., Y. S. Liu and T. L. Low (1979). "Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain." J Biol Chem **254**(8): 2865-2874.
- Radic, M. Z. and M. Zouali (1996). "Receptor editing, immune diversification, and self-tolerance." Immunity **5**(6): 505-511.
- Rakoff-Nahoum, S., J. Paglino, F. Eslami-Varzaneh, S. Edberg and R. Medzhitov (2004). "Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis." Cell **118**(2): 229-241.
- Ramsden, D. A. and M. Gellert (1995). "Formation and resolution of double-strand break intermediates in V(D)J rearrangement." Genes Dev **9**(19): 2409-2420.
- Ramsden, D. A., J. F. McBlane, D. C. van Gent and M. Gellert (1996). "Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage." Embo j **15**(12): 3197-3206.

- Rao, S. P., J. M. Riggs, D. F. Friedman, M. S. Scully, T. W. LeBien and L. E. Silberstein (1999). "Biased VH gene usage in early lineage human B cells: evidence for preferential Ig gene rearrangement in the absence of selection." J Immunol **163**(5): 2732-2740.
- Reaban, M. E. and J. A. Griffin (1990). "Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region." Nature **348**(6299): 342-344.
- Reth, M. (1989). "Antigen receptor tail clue." Nature **338**(6214): 383-384.
- Reth, M., J. Wienands, T. Tsubata and J. Hombach (1991). "Identification of components of the B cell antigen receptor complex." Adv Exp Med Biol **292**: 207-214.
- Retter, M. W. and D. Nemazee (1998). "Receptor editing occurs frequently during normal B cell development." J Exp Med **188**(7): 1231-1238.
- Reynaud, D., I. A. Demarco, K. L. Reddy, H. Schjerven, E. Bertolino, Z. Chen, S. T. Smale, S. Winandy and H. Singh (2008). "Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros." Nat Immunol **9**(8): 927-936.
- Rezanka, L. J., J. J. Kenny and D. L. Longo (2005). "Dual isotype expressing B cells [kappa(+)/lambda(+)] arise during the ontogeny of B cells in the bone marrow of normal nontransgenic mice." Cell Immunol **238**(1): 38-48.
- Rhee, K. J., P. Sethupathi, A. Driks, D. K. Lanning and K. L. Knight (2004). "Role of commensal bacteria in development of gut-associated lymphoid tissues and preimmune antibody repertoire." J Immunol **172**(2): 1118-1124.
- Ribeiro de Almeida, C., R. Stadhouders, M. J. de Bruijn, I. M. Bergen, S. Thongjuea, B. Lenhard, W. van Ijcken, F. Grosveld, N. Galjart, E. Soler and R. W. Hendriks (2011). "The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus." Immunity **35**(4): 501-513.
- Ribeiro de Almeida, C., R. Stadhouders, S. Thongjuea, E. Soler and R. W. Hendriks (2012). "DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation." Blood **119**(26): 6209-6218.

- Richl, P., U. Stern, P. E. Lipsky and H. J. Girschick (2008). "The lambda gene immunoglobulin repertoire of human neonatal B cells." Mol Immunol **45**(2): 320-327.
- Roark, J. H., A. Bui, K. A. Nguyen, L. Mandik and J. Erikson (1997). "Persistence of functionally compromised anti-double-stranded DNA B cells in the periphery of non-autoimmune mice." Int Immunol **9**(11): 1615-1626.
- Rodgers, K. K., Z. Bu, K. G. Fleming, D. G. Schatz, D. M. Engelman and J. E. Coleman (1996). "A zinc-binding domain involved in the dimerization of RAG1." J Mol Biol **260**(1): 70-84.
- Rogers, J., P. Early, C. Carter, K. Calame, M. Bond, L. Hood and R. Wall (1980). "Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain." Cell **20**(2): 303-312.
- Roldan, E., M. Fuxa, W. Chong, D. Martinez, M. Novatchkova, M. Busslinger and J. A. Skok (2005). "Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene." Nat Immunol **6**(1): 31-41.
- Rooney, S., F. W. Alt, J. Sekiguchi and J. P. Manis (2005). "Artemis-independent functions of DNA-dependent protein kinase in Ig heavy chain class switch recombination and development." Proc Natl Acad Sci U S A **102**(7): 2471-2475.
- Ryser, J. E. and P. Vassalli (1974). "Mouse bone marrow lymphocytes and their differentiation." J Immunol **113**(3): 719-728.
- Salzer, U., H. M. Chapel, A. D. Webster, Q. Pan-Hammarstrom, A. Schmitt-Graeff, M. Schlesier, H. H. Peter, J. K. Rockstroh, P. Schneider, A. A. Schaffer, L. Hammarstrom and B. Grimbacher (2005). "Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans." Nat Genet **37**(8): 820-828.
- Samuels, J., Y. S. Ng, C. Coupillaud, D. Paget and E. Meffre (2005). "Impaired early B cell tolerance in patients with rheumatoid arthritis." J Exp Med **201**(10): 1659-1667.
- Santos, P., F. Arumemi, K. S. Park, L. Borghesi and C. Milcarek (2011). "Transcriptional and epigenetic regulation of B cell development." Immunol Res **50**(2-3): 105-112.

- Schamel, W. W. and M. Reth (2000). "Monomeric and oligomeric complexes of the B cell antigen receptor." Immunity **13**(1): 5-14.
- Schatz, D. G. and Y. Ji (2011). "Recombination centres and the orchestration of V(D)J recombination." Nat Rev Immunol **11**(4): 251-263.
- Schatz, D. G., M. A. Oettinger and D. Baltimore (1989). "The V(D)J recombination activating gene, RAG-1." Cell **59**(6): 1035-1048.
- Schatz, D. G. and P. C. Swanson (2011). "V(D)J recombination: mechanisms of initiation." Annu Rev Genet **45**: 167-202.
- Schelonka, R. L., E. Szymanska, A. M. Vale, Y. Zhuang, G. L. Gartland and H. W. Schroeder, Jr. (2010). "DH and JH usage in murine fetal liver mirrors that of human fetal liver." Immunogenetics **62**(10): 653-666.
- Schoettler, N., D. Ni and M. Weigert (2012). "B cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients." Mol Immunol **51**(3-4): 273-282.
- Schug, T. T., D. C. Berry, N. S. Shaw, S. N. Travis and N. Noy (2007). "Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors." Cell **129**(4): 723-733.
- Sela-Culang, I., V. Kunik and Y. Ofran (2013). "The structural basis of antibody-antigen recognition." Front Immunol **4**: 302.
- Shapiro, A. M., M. S. Schlissel, D. Baltimore and A. L. DeFranco (1993). "Stimulation of kappa light-chain gene rearrangement by the immunoglobulin mu heavy chain in a pre-B-cell line." Mol Cell Biol **13**(9): 5679-5690.
- Shikina, T., T. Hiroi, K. Iwatani, M. H. Jang, S. Fukuyama, M. Tamura, T. Kubo, H. Ishikawa and H. Kiyono (2004). "IgA class switch occurs in the organized nasopharynx- and gut-associated lymphoid tissue, but not in the diffuse lamina propria of airways and gut." J Immunol **172**(10): 6259-6264.
- Shockett, P. and J. Stavnezer (1991). "Effect of cytokines on switching to IgA and alpha germline transcripts in the B lymphoma I.29 mu. Transforming growth factor-beta activates transcription of the unrearranged C alpha gene." J Immunol **147**(12): 4374-4383.

- Shulman, Z., A. D. Gitlin, J. S. Weinstein, B. Lainez, E. Esplugues, R. A. Flavell, J. E. Craft and M. C. Nussenzweig (2014). "Dynamic signaling by T follicular helper cells during germinal center B cell selection." Science **345**(6200): 1058-1062.
- Siminovitch, K. A., A. Bakhshi, P. Goldman and S. J. Korsmeyer (1985). "A uniform deleting element mediates the loss of kappa genes in human B cells." Nature **316**(6025): 260-262.
- Sims, G. P., R. Ettinger, Y. Shirota, C. H. Yarboro, G. G. Illei and P. E. Lipsky (2005). "Identification and characterization of circulating human transitional B cells." Blood **105**(11): 4390-4398.
- Spencer, J., F. Barone and D. Dunn-Walters (2009). "Generation of Immunoglobulin diversity in human gut-associated lymphoid tissue." Semin Immunol **21**(3): 139-146.
- Spencer, J., L. S. Klavinskis and L. D. Fraser (2012). "The human intestinal IgA response; burning questions." Front Immunol **3**: 108.
- Spicuglia, S., D. M. Franchini and P. Ferrier (2006). "Regulation of V(D)J recombination." Curr Opin Immunol **18**(2): 158-163.
- Spits, H., B. Blom, A. C. Jaleco, K. Weijer, M. C. Verschuren, J. J. van Dongen, M. H. Heemskerk and P. C. Res (1998). "Early stages in the development of human T, natural killer and thymic dendritic cells." Immunol Rev **165**: 75-86.
- Stamatopoulos, K., C. Belessi, A. Hadzidimitriou, T. Smilevska, E. Kalagiakou, K. Hatzi, N. Stavroyianni, A. Athanasiadou, A. Tsompanakou, T. Papadaki, G. Kokkini, G. Paterakis, R. Saloum, N. Laoutaris, A. Anagnostopoulos and A. Fassas (2005). "Immunoglobulin light chain repertoire in chronic lymphocytic leukemia." Blood **106**(10): 3575-3583.
- Stavnezer, J. (1996). "Antibody class switching." Adv Immunol **61**: 79-146.
- Stavnezer, J., J. E. Guikema and C. E. Schrader (2008). "Mechanism and regulation of class switch recombination." Annu Rev Immunol **26**: 261-292.
- Steinsbo, O., C. J. Henry Dunand, M. Huang, L. Mesin, M. Salgado-Ferrer, K. E. Lundin, J. Jahnsen, P. C. Wilson and L. M. Sollid (2014). "Restricted VH/VL usage and limited mutations in gluten-specific IgA of coeliac disease lesion plasma cells." Nat Commun **5**: 4041.

- Su, T. T. and D. J. Rawlings (2002). "Transitional B lymphocyte subsets operate as distinct checkpoints in murine splenic B cell development." J Immunol **168**(5): 2101-2110.
- Su, W., L. Boursier, A. Padala, J. D. Sanderson and J. Spencer (2004). "Biases in Ig Light Chain Rearrangements in Human Intestinal Plasma Cells." The Journal of Immunology **172**(4): 2360-2366.
- Su, W., J. N. Gordon, F. Barone, L. Boursier, W. Turnbull, S. Mendis, D. K. Dunn-Walters and J. Spencer (2008). "Lambda Light Chain Revision in the Human Intestinal IgA Response." The Journal of Immunology **181**(2): 1264-1271.
- Suzuki, K. and S. Fagarasan (2009). "Diverse regulatory pathways for IgA synthesis in the gut." Mucosal Immunol **2**(6): 468-471.
- Suzuki, K., B. Meek, Y. Doi, M. Muramatsu, T. Chiba, T. Honjo and S. Fagarasan (2004). "Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut." Proc Natl Acad Sci U S A **101**(7): 1981-1986.
- Swaminathan, S., C. Duy and M. Muschen (2014). "BACH2-BCL6 balance regulates selection at the pre-B cell receptor checkpoint." Trends Immunol **35**(3): 131-137.
- Swanson, P. C. (2004). "The bounty of RAGs: recombination signal complexes and reaction outcomes." Immunol Rev **200**: 90-114.
- Swanson, P. C., S. Kumar and P. Raval (2009). "Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes." Adv Exp Med Biol **650**: 1-15.
- Taccioli, G. E., G. Rathbun, E. Oltz, T. Stamato, P. A. Jeggo and F. W. Alt (1993). "Impairment of V(D)J recombination in double-strand break repair mutants." Science **260**(5105): 207-210.
- Tachibana, H., H. Haruta and K. Yamada (1999). "Light chain shifting: identification of a human plasma cell line actively undergoing light chain replacement." Blood **93**(1): 198-207.
- ten Boekel, E., F. Melchers and A. Rolink (1995). "The status of Ig loci rearrangements in single cells from different stages of B cell development." Int Immunol **7**(6): 1013-1019.

- ten Boekel, E., F. Melchers and A. G. Rolink (1998). "Precursor B cells showing H chain allelic inclusion display allelic exclusion at the level of pre-B cell receptor surface expression." Immunity **8**(2): 199-207.
- Tiegs, S. L., D. M. Russell and D. Nemazee (1993). "Receptor editing in self-reactive bone marrow B cells." J Exp Med **177**(4): 1009-1020.
- Tomana, M., W. Niedermeier, J. Mestecky and F. Skvaril (1976). "The differences in carbohydrate composition between the subclasses of IgA immunoglobulins." Immunochemistry **13**(4): 325-328.
- Tomasi, T. B., Jr., E. M. Tan, A. Solomon and R. A. Prendergast (1965). "CHARACTERISTICS OF AN IMMUNE SYSTEM COMMON TO CERTAIN EXTERNAL SECRETIONS." J Exp Med **121**: 101-124.
- Tomlinson, I. M., J. P. Cox, E. Gherardi, A. M. Lesk and C. Chothia (1995). "The structural repertoire of the human V kappa domain." Embo j **14**(18): 4628-4638.
- Tsuji, M., N. Komatsu, S. Kawamoto, K. Suzuki, O. Kanagawa, T. Honjo, S. Hori and S. Fagarasan (2009). "Preferential generation of follicular B helper T cells from Foxp3+ T cells in gut Peyer's patches." Science **323**(5920): 1488-1492.
- Vaerman, J. P., A. E. Langendries, D. A. Giffroy, C. S. Kaetzel, C. M. Fiani, I. Moro, P. Brandtzaeg and K. Kobayashi (1998). "Antibody against the human J chain inhibits polymeric Ig receptor-mediated biliary and epithelial transport of human polymeric IgA." Eur J Immunol **28**(1): 171-182.
- van der Burg, M., T. Tumkaya, M. Boerma, S. de Bruin-Versteeg, A. W. Langerak and J. J. van Dongen (2001). "Ordered recombination of immunoglobulin light chain genes occurs at the IGK locus but seems less strict at the IGL locus." Blood **97**(4): 1001-1008.
- van Gent, D. C., D. A. Ramsden and M. Gellert (1996). "The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination." Cell **85**(1): 107-113.
- van Zelm, M. C., M. van der Burg, D. de Ridder, B. H. Barendregt, E. F. de Haas, M. J. Reinders, A. C. Lankester, T. Revesz, F. J. Staal and J. J. van Dongen (2005). "Ig gene rearrangement steps are initiated in early human precursor B cell subsets and

correlate with specific transcription factor expression." J Immunol **175**(9): 5912-5922.

- Vela, J. L., D. Ait-Azzouzene, B. H. Duong, T. Ota and D. Nemazee (2008). "Rearrangement of mouse immunoglobulin kappa deleting element recombining sequence promotes immune tolerance and lambda B cell production." Immunity **28**(2): 161-170.
- Venigalla, R. K., V. A. McGuire, R. Clarke, J. C. Patterson-Kane, A. Najafov, R. Toth, P. C. McCarthy, F. Simeons, L. Stojanovski and J. S. Arthur (2013). "PDK1 regulates VDJ recombination, cell-cycle exit and survival during B-cell development." Embo j **32**(7): 1008-1022.
- Vettermann, C. and M. S. Schlissel (2010). "Allelic exclusion of immunoglobulin genes: models and mechanisms." Immunol Rev **237**(1): 22-42.
- Victora, G. D. and M. C. Nussenzweig (2012). "Germinal centers." Annu Rev Immunol **30**: 429-457.
- Victora, G. D., T. A. Schwickert, D. R. Fooksman, A. O. Kamphorst, M. Meyer-Hermann, M. L. Dustin and M. C. Nussenzweig (2010). "Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter." Cell **143**(4): 592-605.
- Volpe, J. M. and T. B. Kepler (2008). "Large-scale analysis of human heavy chain V(D)J recombination patterns." Immunome Res **4**: 3.
- von Muenchow, L., C. Engdahl, K. Karjalainen and A. G. Rolink (2014). "The selection of mature B cells is critically dependent on the expression level of the co-receptor CD19." Immunol Lett.
- Vossenkamper, A., P. A. Blair, N. Safinia, L. D. Fraser, L. Das, T. J. Sanders, A. J. Stagg, J. D. Sanderson, K. Taylor, F. Chang, L. M. Choong, D. P. D'Cruz, T. T. Macdonald, G. Lombardi and J. Spencer (2013). "A role for gut-associated lymphoid tissue in shaping the human B cell repertoire." J Exp Med **210**(9): 1665-1674.
- Wardemann, H., S. Yurasov, A. Schaefer, J. W. Young, E. Meffre and M. C. Nussenzweig (2003). "Predominant autoantibody production by early human B cell precursors." Science **301**(5638): 1374-1377.

- Warnock, R. A., J. J. Campbell, M. E. Dorf, A. Matsuzawa, L. M. McEvoy and E. C. Butcher (2000). "The role of chemokines in the microenvironmental control of T versus B cell arrest in Peyer's patch high endothelial venules." J Exp Med **191**(1): 77-88.
- Watson, C. T. and F. Breden (2012). "The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease." Genes Immun **13**(5): 363-373.
- Weber, J. C., G. Blaison, T. Martin, A. M. Knapp and J. L. Pasquali (1994). "Evidence that the V kappa III gene usage is nonstochastic in both adult and newborn peripheral B cells and that peripheral CD5+ adult B cells are oligoclonal." J Clin Invest **93**(5): 2093-2105.
- Wen, L., J. Brill-Dashoff, S. A. Shinton, M. Asano, R. R. Hardy and K. Hayakawa (2005). "Evidence of marginal-zone B cell-positive selection in spleen." Immunity **23**(3): 297-308.
- Widhopf, G. F., 2nd, C. J. Goldberg, T. L. Toy, L. Z. Rassenti, W. G. Wierda, J. C. Byrd, M. J. Keating, J. G. Gribben, K. R. Rai and T. J. Kipps (2008). "Nonstochastic pairing of immunoglobulin heavy and light chains expressed by chronic lymphocytic leukemia B cells is predicated on the heavy chain CDR3." Blood **111**(6): 3137-3144.
- Wirths, S. and A. Lanzavecchia (2005). "ABCB1 transporter discriminates human resting naive B cells from cycling transitional and memory B cells." Eur J Immunol **35**(12): 3433-3441.
- Wolf, G. (2007). "Identification of a membrane receptor for retinol-binding protein functioning in the cellular uptake of retinol." Nutr Rev **65**(8 Pt 1): 385-388.
- Woodward, E. J. and J. W. Thomas (2005). "Multiple germline kappa light chains generate anti-insulin B cells in nonobese diabetic mice." J Immunol **175**(2): 1073-1079.
- Woof, J. M. and J. Mestecky (2005). "Mucosal immunoglobulins." Immunol Rev **206**: 64-82.

- Wotherspoon, A. C., C. Ortiz-Hidalgo, M. R. Falzon and P. G. Isaacson (1991). "Helicobacter pylori-associated gastritis and primary B-cell gastric lymphoma." Lancet **338**(8776): 1175-1176.
- Wu, T. T. and E. A. Kabat (1970). "An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity." J Exp Med **132**(2): 211-250.
- Wu, Y. C., D. Kipling, H. S. Leong, V. Martin, A. A. Ademokun and D. K. Dunn-Walters (2010). "High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations." Blood **116**(7): 1070-1078.
- Yancopoulos, G. D. and F. W. Alt (1985). "Developmentally controlled and tissue-specific expression of unrearranged VH gene segments." Cell **40**(2): 271-281.
- Yancopoulos, G. D., S. V. Desiderio, M. Paskind, J. F. Kearney, D. Baltimore and F. W. Alt (1984). "Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines." Nature **311**(5988): 727-733.
- Yurasov, S., T. Tiller, M. Tsuiji, K. Velinzon, V. Pascual, H. Wardemann and M. C. Nussenzweig (2006). "Persistent expression of autoantibodies in SLE patients in remission." J Exp Med **203**(10): 2255-2261.
- Yurasov, S., H. Wardemann, J. Hammersen, M. Tsuiji, E. Meffre, V. Pascual and M. C. Nussenzweig (2005). "Defective B cell tolerance checkpoints in systemic lupus erythematosus." J Exp Med **201**(5): 703-711.
- Zheng, N. Y., K. Wilson, X. Wang, A. Boston, G. Kolar, S. M. Jackson, Y. J. Liu, V. Pascual, J. D. Capra and P. C. Wilson (2004). "Human immunoglobulin selection associated with class switch and possible tolerogenic origins for C delta class-switched B cells." J Clin Invest **113**(8): 1188-1201.
- Zotos, D. and D. M. Tarlinton (2012). "Determining germinal centre B cell fate." Trends Immunol **33**(6): 281-288.

Molecular Immunology 65 (2015) 215–223

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Immunoglobulin kappa variable region gene selection during early human B cell development in health and systemic lupus erythematosus

Verena Hehle[a,1], Louise D. Fraser[a,1], Romeeza Tahir[a], David Kipling[b], Yu-Chang Wu[a], Pamela M.K. Lutalo[a,c], John Cason[a], LeeMeng Choong[c], David P. D'Cruz[c], Andrew P. Cope[d], Deborah K. Dunn-Walters[a], Jo Spencer[a,*]

[a] Department of Immunobiology, King's College London, London, UK
[b] School of Medicine, Cardiff University, Cardiff, UK
[c] Louise Coote Lupus Unit Guy's and St Thomas' NHS Trust, London, UK
[d] Academic Department of Rheumatology, King's College London, London, UK

ARTICLE INFO

Article history:
Received 1 December 2014
Received in revised form 13 January 2015
Accepted 15 January 2015
Available online 17 February 2015

Keywords:
B cell
Immunoglobulin
Kappa light chains
Repertoire
Selection

ABSTRACT

The unique specificity of the B cell receptor is generated by an ordered sequence of gene rearrangement events. Once IGH genes have rearranged, rearrangement at the *IGK* locus is initiated followed by the *IGL* locus if functional *IGK* rearrangement is not achieved. Receptor specificity can subsequently be altered by secondary light chain editing based on the features of the heavy and light chain combination. The final profile of expressed genes is not random and biases in this profile are associated with several autoimmune diseases. However, how and when biases are created is not known.

To increase our understanding of the processes of selection and editing of *IGK* rearrangements, we compared four groups of rearrangements of IGK acquired by next generation sequencing. First, expressed rearrangements of *IGK* from cDNA of IGK expressing B cells. Second, productive rearrangements of *IGK* from DNA of the same kappa expressing B cells. Third, non-productive rearrangements of *IGK* from DNA of IGK and IGL expressing B cells, and fourth productively rearranged *IGK* from DNA of IGL expressing B cells. The latter group would have been rejected during B cell development in favour of rearrangement at the *IGL* locus and are therefore selected against.

We saw evidence that rearranged IGK segments can be selected at a checkpoint where the decision to rearrange the IGL locus is made. In addition, our data suggest that mechanisms regulating the expression or not of IGK rearrangements may also contribute to repertoire development and also that this latter component of the selection process is defective in SLE.

1. Introduction

The unique specificity of individual B cells for antigen is determined by the sequences of the expressed *IGH* and *IGK* or *IGL* light chain genes. The sequences are created by an ordered succession of rearrangement events starting with rearrangement of variable (V), diversity (D) and joining (J) segments of *IGH*. Successful expression of the pre-B cell receptor excludes recombination events on the second *IGH* allele and also activates enhancers that initiate rearrangement of the *IGK* locus. If functional rearrangement is not achieved on either allele at the *IGK* locus, then *IGK* is inactivated by deletional rearrangement with kappa deleting element (KDE) and gene rearrangement proceeds at the *IGL* locus (Hardy and Hayakawa, 2001; Santos et al., 2011).

The final shape and specificity of the BCR can potentially be altered many times during early B cell development by secondary editing rearrangements of segments upstream and downstream of an original *V-J* rearrangement of *IGK or IGL* (Wardemann et al., 2003; Yurasov et al., 2005b). Receptor editing may occur if the receptor as

* Corresponding author at: Department of Immunobiology, King's College London, Guy's Campus, St Thomas' St, London, SE1 9RT, UK. Tel.: +44 207 188 1614; fax: +44 207 118 3385.
E-mail addresses: verena.hehle@kcl.ac.uk (V. Hehle), louisefraser13@googlemail.com (L.D. Fraser), romeeza.tahir@kcl.ac.uk (R. Tahir), kiplingd@cardiff.ac.uk (D. Kipling), yu-chang.wu@kcl.ac.uk (Y.-C. Wu), pamela.lutalo@kcl.ac.uk (P.M.K. Lutalo), john.cason@kcl.ac.uk (J. Cason), leemeng.choong@gstt.nhs.uk (L. Choong), david.d'cruz@kcl.ac.uk (D.P. D'Cruz), Andrew.cope@kcl.ac.uk (A.P. Cope), deborah.dunn-walters@kcl.ac.uk (D.K. Dunn-Walters), jo.spencer@kcl.ac.uk (J. Spencer).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.molimm.2015.01.017
0161-5890/

a whole is autoreactive, and development of autoimmune diseases may result from a failure to edit the BCR effectively (Panigrahi et al., 2008). Autoimmune diseases are frequently associated with biases in light chain gene segment usage, e.g., there is an over representation of *IGKV4*-1 in systemic lupus erythematosus (SLE) (Dorner et al., 1998; Woodward and Thomas, 2005), celiac disease (Steinsbø et al., 2014) and type 1 diabetes (Woodward and Thomas, 2005).

The *IGK* locus includes gene segments in forward and reverse orientations, relative to the downstream gene segments. For example, the *IGKV4-1* and *IGKV5*-2 genes most proximal to the J segments are in reverse transcriptional orientation compared to the other proximal *IGKV* genes (Foster et al., 1997; http://www.imgt.org/IMGTrepertoire/; Schoettler et al., 2012). When *IGKV4*-1 or *IGKV5*-2 recombine with *IGKJ* the DNA coils to permit alignment of the recombination signal sequences (RSS). After rearrangement, the intervening DNA between *IGKV* and *IGKJ* is not released as an episomal fragment, but instead the rearrangement is inserted in inverted orientation adjacent to or among the J segments. This rearrangement can be retained on the chromosome even after secondary editing rearrangements. It is therefore possible that KDE rearrangements to the intronic recombinant signal sequences (RSS) may not remove the edited gene rearrangements of *IGK* that have been selected against and cells may accumulate more than one unused rearrangement on a single chromosome. Therefore B cells can accumulate rearrangements that have been excluded from the functional repertoire. It is also possible that 'same orientation' secondary *IGK* rearrangements that excise existing rearrangements as episomes may still be amplifiable by PCR. This would be apparent as a high relative frequency of productive DNA rearrangements of an *IGK* gene segment compared to expressed sequences in cDNA and can be considered to be a consequence of accumulation of rearrangements in DNA that have been selected against during B cell development. Therefore, rearrangements of V genes that occur relatively more frequently in DNA compared to cDNA from the same cells may have been selected against.

Our aim was to analyse *IGKV* gene selection in humans and how this might be altered in SLE. To do this we have studied four groups of *IGK* sequences (1–4 in Fig. 1). The first group of sequences (1) was PCR amplified from cDNA of IGK expressing B cells. These are considered to be the gold standard indication of actual *IGKV* gene expression in the B cell population in our study. The second group of sequences (2) was productive *IGK* gene rearrangements from DNA of the same, sorted IGK expressing mature naïve B cell subset used for cDNA preparation. The third group of sequences studied (3) was the DNA rearrangements that were unproductive due to being out of frame or having a stop codon. The B cells expressing these genetic rearrangements of the *IGK* locus are not selected since they do not encode any protein. Their profile therefore reflects the intrinsic biases in the efficiency of rearrangement of *IGKV* gene segments determined, e.g., by transcriptional activity and RSS sequence and not the properties of the *IGKV* segment encoded (Foster et al., 1997). The fourth group (4) of sequences studied was productive DNA rearrangements of *IGK* from sorted IGL expressing B cells from the same donors as the IGK expressing cells. Since the B cells can change from rearrangement of *IGK* to rearrangement of the *IGL* locus at the pre-B stage, this group of productive rearrangements has been rejected during B cell development (Brauninger et al., 2001). This is summarised in Fig. 1.

By comparing these four groups of sequences we observed that some *IGKV* segments are consistently selected either for or against expression during early B cell development. We therefore compared the rearrangement profile of the four groups of sequences in six healthy controls and three patients with SLE. Whereas the overall imprint on repertoire selection at the DNA level is similar including the switch from *IGK* to *IGL* gene rearrangements, there are differences between health and SLE when comparing DNA and cDNA from IGK expressing B cells. This implies that mechanisms of gene silencing rather than detection of self-reactivity per se may be defective in SLE.

2. Materials and methods

2.1. B-cell isolation and cell sorting of naïve B cells

Peripheral blood mononuclear cells were isolated from six healthy donors, and three SLE patients using Ficoll-Paque Plus (GE

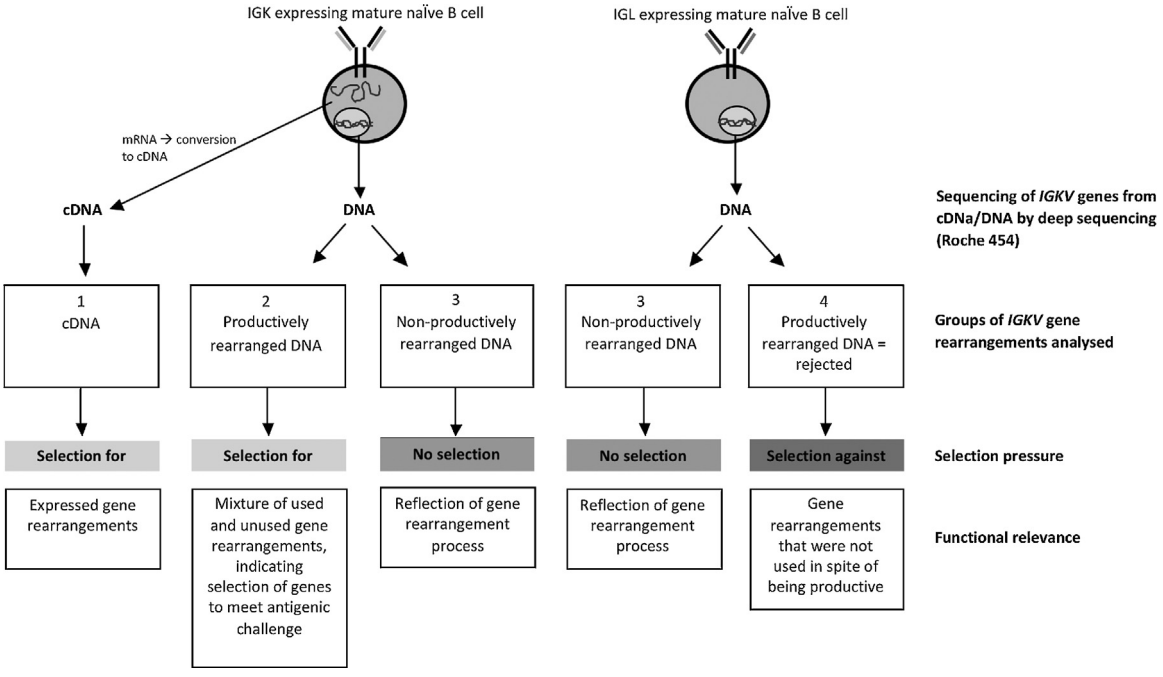


Fig. 1. Flow chart showing groups if IGKV gene rearrangements analysed and functional relevance. Genomic DNA and RNA were extracted from IGK and IGL surface expressing mature naïve B cells. Four groups of sequences are depicted and functional relevance of different subsets of IGKV gene rearrangements analysed are illustrated.

Healthcare). Samples from three ANA+ patients with SLE were collected through the King's College London Infectious Diseases Biobank and written consent was obtained in accordance with Guy's Hospital research ethics committee. The SLE patients were females aged 40 (Afro-Carribean, 12 years history of SLE, treated with mycophenolate mofetil and hydroxychloroquine, nephritis), 59 (Caucasian, 4 years history of SLE, treated with Leflunomide, arthritis) and 32 (Caucasian, 3 years history of SLE, treated with prednisolone and hydroxychloroquine, arthritis).

PBMCs were stained with CD19-PerCP-Cy5.5 (BD PharMingen), kappa light chain-APC-H7 (BD PharMingen) and lambda light chain-Pacific Blue (Biolegend), CD27 APC (BD PharMingen), IgD PE-Cy7 (BD Biosciences), and LIVE/DEAD Fixable Aqua (Life Technologies) for 15 min at 4 °C before cell sorting on the FACSAria machine (BD Biosciences PharMingen.) IGK expressing B mature naïve B cells ($CD19^+$ IgD^+ $CD27^-$ IGK^+) and IGL expressing B mature naïve B cells ($CD19^+$ IgD^+ $CD27^-$ IGL^+) were collected separately into 180 μl of Sort-Lysis RT buffer (SLyRT) (Wu et al., 2010). B cells for sequencing of genomic DNA were collected in RPMI. Genomic DNA was extracted using QIAGEN DNasey Blood and Tissue Kit (see suppliers instructions). Unlike healthy control samples, SLE cases did not always show clear segregation of IGK- and IGL-expressing cells due to apparent light chain inclusion in some cases (Fraser et al., unpublished). The cases used in this study were those where discrete IGK- and IGL-expressing populations could be isolated.

2.2. cDNA synthesis and Ig PCR

To synthesise cDNA, 500U SuperScript III reverse transcriptase (RT; Invitrogen) in 20 μl were added to 180 μl of cells in SLyRT buffer that contained 300 ng pd(N)6 (Qiagen) in 60 μl. RT was performed at 42 °C (10 min), 25 °C (10 min), 50 °C (60 min) and 72 °C (15 min). Ig genes were amplified using semi-nested polymerase chain reaction (PCR). In PCR1, a 25-μL reaction mix contained 6.25 μL of cDNA, 0.625 U Phusion DNA polymerase (NEB), 200 μM each dNTPs, 41.75 nM each upstream *IGK* primers, 250 nM downstream *IGKC* primer in 1× reaction buffer. After a hot start at 98 °C (for 30 s, hold at 50 °C) Phusion DNA polymerase was added, followed by 15 cycles of 98 °C (10 s), 58 °C (15 s), 72 °C (30 s), and 1 cycle of 72 °C (5 min). PCR2 was then used to amplify from PCR1 products using primers having 10-base multiplex-identifier (MID) tails. MID tags enabled nine different samples to be pooled into one sequencing sample and individual experimental samples were later separated by sequence analysis of the MID tags. Twenty microlitres of PCR2 reaction mix contained 2 μl of PCR1 products, 0.5 U Phusion DNA polymerase, 200 μM each dNTPs, 41.75 nM each upstream MID: IGK 1–6 primers, 250 nM *IGKC* downstream primer in 1× reaction buffer. PCR2 was performed at: 98 °C for (30 s), 20 cycles of 98 °C (10 s), 58 °C (15 s), 72 °C (30 s), and 1 cycle of 72 °C (5 min). Primer sequences are given in Supplementary Tables S1 and S2.

2.3. Preparation of MID PCR products for next generation sequencing

To produce sufficient DNA for sequencing while minimising PCR amplification, repeats of PCR1 (eight sample aliquots) and PCR2 (two sample aliquots from each of the eight PCR products) were performed for each individual experimental sample. PCR Primers were removed from the 16 pooled products by gel electrophoresis and using QIAquick Gel Purification Kit. Samples to be pooled for sequencing were mixed in equal quantities and concentrated using QIAquick PCR Purification Kit (QIAGEN) before sequencing on the Roche 454 Titanium Sequencer (Agowa GmBH). Accuracy of the method as a whole was determined using analysis of results from a control Ig gene and was <1 error per 300 bp, or 1 per 1300 bp if indels (a known issue with the sequencing platform) were discounted (18).

2.4. Sequence analysis

Sequences were assigned to the corresponding samples based on the terminal MID sequence. Sequences that contained a second MID sequence that was either different or located internally, or sequences that had no MID tag were excluded. A series of stringent quality control criteria were applied to exclude biologically implausible sequences (Wu et al., 2010). *IGKV-C* rearrangements amplified from cDNA were only accepted as biologically plausible if they were over 415 nucleotides in length and *IGKV-J* rearrangements amplified from the DNA of IGK- or IGL-expressing B cells were accepted providing they were over 290 nucleotides in length. Sequences that had *IGK* with either *IGKJ* or *IGKC* were rejected, as were sequences with multiple IGV gene primer motifs. Clonally related sequences were identified by sorting based on their CDR3 nucleic acid sequences with *IGK* and *IGKJ* gene use as secondary identifiers and only one was included in the data.

Analyses were performed in Excel (Microsoft). Proportions were compared using χ^2 tests with Bonferroni correction for multiple comparisons and with Student's *t*-test of the mean values with GraphPad Prism.

Data are expressed as frequencies throughout the manuscript. The frequency in each case is the percentage of the IGK V or J segment of the total number of sequences in the group analysed.

3. Results

Clustering of complementarity determining region 3 sequences identified a total of 20,160 unique rearrangements of *IGKV- J* for analysis in this study. Individual donors contributed different numbers of sequences to the total as indicated (Table 1). Of note, and consistent with other studies (Brauninger et al., 2001; Goossens et al., 2001), approximately 25% of rearrangements of sequenced *IGK* segments from IGL expressing B cells were productively rearranged. These gene rearrangements have been selected against during B cell development and are referred to as rejected productive DNA rearrangements (Fig. 1).

3.1. Comparison frequencies of rearrangements involving IGKV genes in cDNA and DNA from IGK and IGL sorted B cells from six healthy donors

Initially, the profile of *IGKV* gene rearrangements in cDNA and DNA of IGK expressing B cells from each of the six healthy donors was determined when sequences from all donors were pooled (Fig. 2A) or analysed individually (Figure S1). There was a relatively greater frequency of rearrangements of genes from the proximal cluster closer to the constant region segments compared to the distal cluster. The rearranged *IGK* genes that were consistently significantly relatively more abundant in DNA than cDNA in the pooled data and consistent in six donors were *IGKV1-37*, *IGKV5-2* and *IGKV4-1*. In contrast rearrangements involving *IGKV1-15* and *IGKV3-11* segments were relatively more abundant in cDNA than DNA.

Next, productively rearranged *IGKV* genes from DNA of IGK and IGL expressing B cells were compared. Genes that had been selected against are those where black bars is taller than grey bars in both figures; i.e. they were more abundant in DNA than cDNA in Fig. 2A and more abundant in IGL than IGK expressing B cells in Fig. 2B. Some genes known to be non-functional showed this selection profile including *IGKV1-37* that is non-functional due to the replacement of cysteine at position 104 in framework region 3 with glycine, which is consistent with our interpretation of the data (Pargent

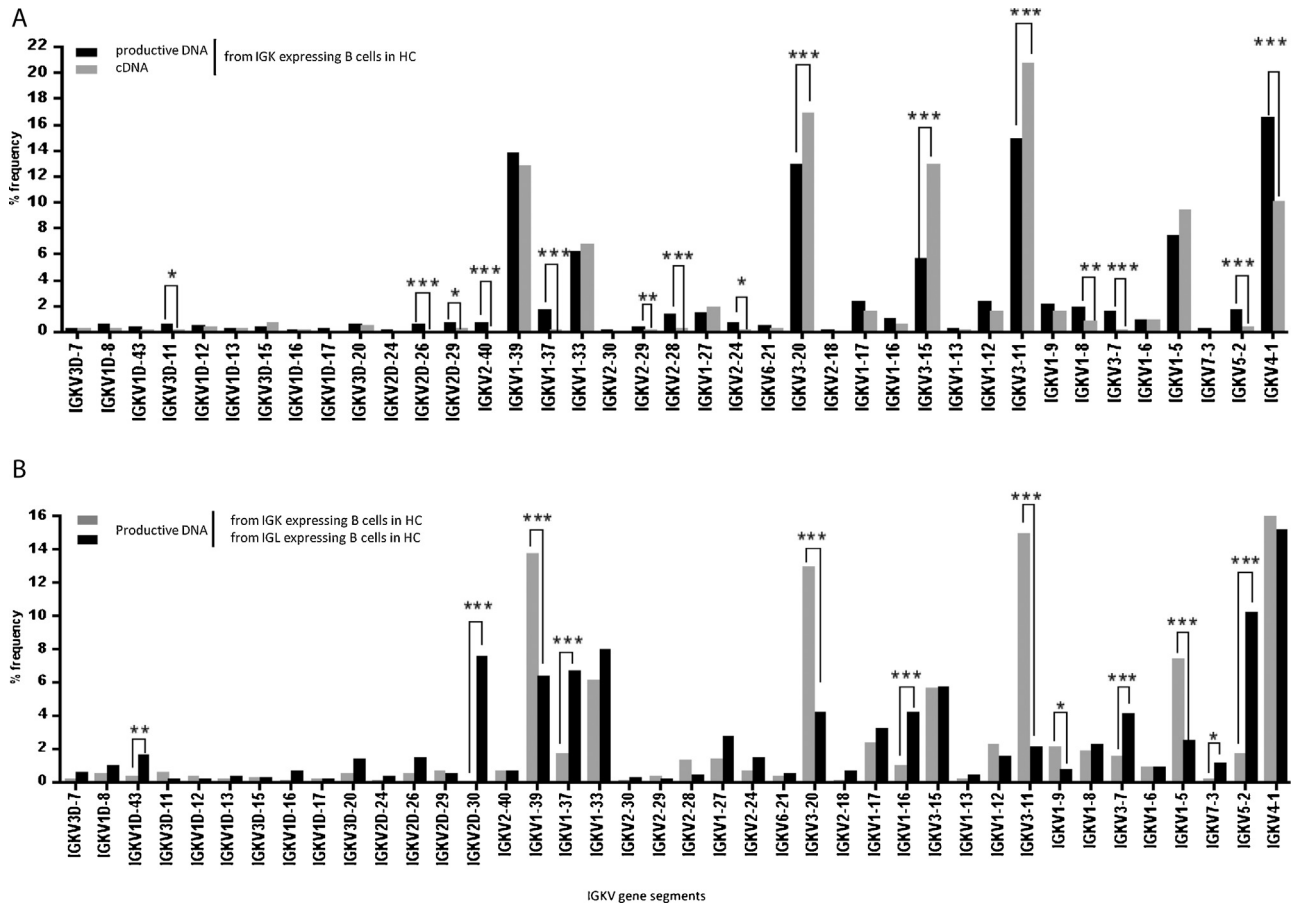


Fig. 2. Comparison of relative frequencies of rearrangements of *IGKV* gene segments with data from six healthy controls. (A) Comparison of PCR amplicons from DNA (black bars) or cDNA (grey bars) of mature naïve B cells expressing IGK. Where the frequency of rearrangements is higher in DNA than cDNA this gene is more abundantly present in DNA than expressed, consistent with retention of unused, edited segments in the cell. (B) Productive DNA rearrangements of *IGKV* gene segments from IGK (grey bar) and IGL (black bar) expressing mature naïve B cells. Where frequency of IGKV gene segments is higher in IGL expressing B cells, these *IGKV* genes have been selected against. Gene segments which are more abundant in IGK that IGL expressing B cells are selected for. Zero values (see Fig. S1) were excluded from this figure. Statistical analysis was performed using χ^2 analysis adjusted using the Bonferroni correction. *p*-Values are indicated as $^{*}p < 0.05$, $^{**}p < 0.005$ and $^{***}p < 0.0005$.

et al., 1991). Based on this definition of segments that are selected against, it is apparent that expression of *IGKV5-2* is strongly selected against during early B cell development so that despite being relatively frequently rearranged they are rare in cDNA.

Genes that had been selected for are those where the grey bars are taller than the black bars in Fig. 2A and B. They were less abundant in DNA compared to cDNA in Fig. 2A and more common in IGK than IGL expressing B cells in Fig. 2B. An example of a gene that appears to be selected for was *IGKV3-11*.

Gene segment *IGKV4-1* is of particular interest. Although it was consistently and significantly selected against by comparison of DNA and cDNA in health (Figs. 2A and S1), there was no evidence of selection against this segment when rejected productive rearrangements of *IGKV* from IGL expressing cells and productive rearrangements of IGK from IGK expressing cells were compared (Fig. 2B).

3.2. Analysis of IGKV gene selection in SLE

We compared *IGKV* sequences from three SLE patients in Fig. 3A and B, applying the same principles already introduced in Fig. 2A and B. Rearrangement frequencies of individual gene segments indicate selection against *IGKV1-37*, *IGKV5-2* in SLE as in health. In general, however, the differences between DNA and cDNA observed in health, Fig. 2A, were not so apparent in SLE in Fig. 3A. For example, there was no difference between the frequencies of the IGKV4-1 gene in Fig. 3A as observed in health in Fig. 2A.

We next focused on the selection profile of individual genes of interest that showed different hallmarks of selection. *IGKV3-11* was selected because it consistently appeared to be favoured. *IGKV1-37* was selected because it is non-functional due to the replacement of cysteine with glycine in framework region 3 and therefore selected against for reasons we can understand (Fig. 4A–D). It is apparent by comparison of data relating to *IGKV5-2* with the 'control' profiles of *IGKV3-11* and *IGKV1-37* that *IGKV5-2* had the profile of a gene that had been selected against during B cell development in health and SLE (Fig. 4C and H). This segment has low frequencies in the cDNA pool compared to DNA, and rearrangements are more abundant in productive rearrangements in DNA from IGL than IGK expressing B cells in both health and SLE. *IGKV1-17* is illustrated because it showed selection against in health, but not in SLE. *IGKV4-1* (Fig. 4E and J) could not be clearly assigned as either selected for or against by any of the criteria that identified selection in other gene segments by analysis of groups of rearrangements amplified from DNA alone. However, frequencies of *IGKV4-1* were consistently and significantly lower in cDNA than in DNA from the same cells in health but interestingly, not in SLE.

Since the pooled data suggested that the lower frequency of *IGKV4-1* in cDNA compared to DNA in health, was not so marked in SLE, the frequencies of *IGKV4-1* in cDNA and DNA in individual blood donors in health and SLE was analysed (Fig. 5). Whilst the

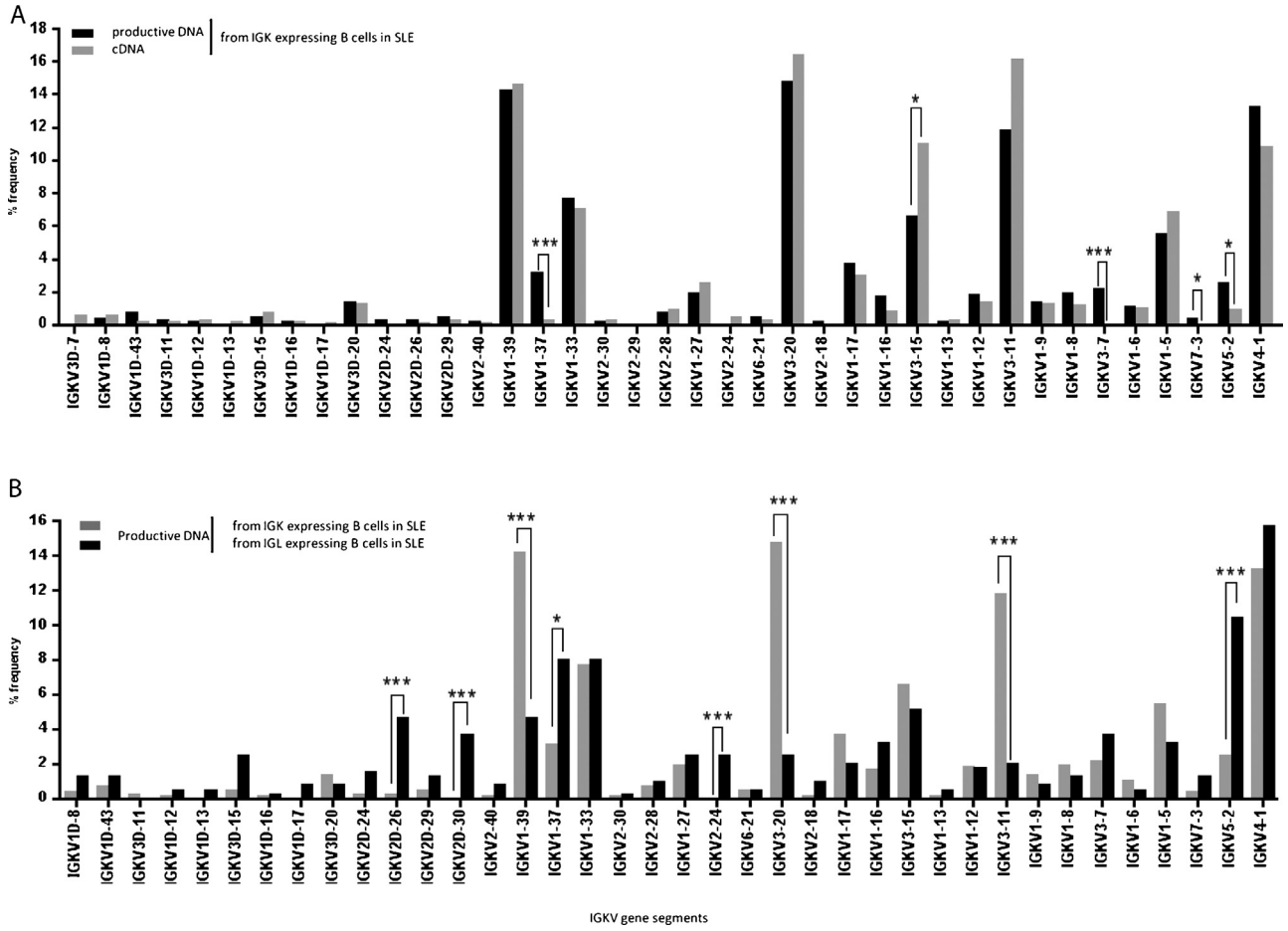


Fig. 3. Comparison of relative frequencies of rearrangements of IGKV gene segments with data from three SLE patients. (A) Comparison of PCR amplicons from DNA (black bars) or cDNA (grey bars) of mature naïve B cells expressing IGK. Where the frequency of rearrangements is higher in DNA than cDNA this gene is more abundantly present in DNA than expressed, consistent with retention of unused, edited segments in the cell. (B) Productive DNA rearrangements of *IGKV* gene segments from IGK (grey bar) and IGL (black bar) expressing mature naïve B cells. Where frequency of IGKV gene segments is higher in IGL expressing B cells, these *IGKV* genes have been selected against. Gene segments which are more abundant in IGK that IGL expressing B cells are selected for. Zero values were excluded from this figure. Statistical analysis was performed using χ^2 analysis adjusted using the Bonferroni correction. *P*-Values are indicated as $^{*}p < 0.05$, $^{**}p < 0.005$ and $^{***}p < 0.0005$.

lower frequency of *IGKV4-1* in cDNA compared to DNA is apparent in all six donors in health and is statistically significant as a population, this was not observed in two of the three patients with SLE.

3.3. Comparison of IGKJ segment usage in health and SLE

We investigated the use of *IGKJ* segments in productively rearranged DNA and cDNA from healthy donors and donors with SLE (Fig. 6). There was a significant tendency to use *IGKJ1* and *IGKJ4* and not to use *IGKJ5* in cDNA compared to DNA in health and SLE. This did not differ between IGKV genes analysed (data not shown). *IGKJ2* was rearranged significantly less in DNA of patients with SLE compared to controls.

4. Discussion

Our study has identified selection biases for and against the expression of individual *IGKV* gene segments during the development of the mature naïve B cell repertoire in healthy individuals and patients with SLE. The frequency of non-productive rearrangements of individual *IGKV* segments identifies the profile of rearrangement biases with no selection (Foster et al., 1997). When *IGKV* segments are selected against, their frequencies are above this threshold in the rejected productive DNA gene rearrangements *of IGKV* genes from IGL expressing B cells. In addition, they are below the threshold in productive DNA from IGK expressing B cells. The frequency of segments that have been selected against is then lowest in the cDNA. Gene segments selected against according to these criteria include *IGKV3-7*, *IGKV1-37* and *IGKV5-2*. Selection against rearrangements of *IGKV1-37* and *IGKV3-7* is expected because *IGKV1-37* is non-functional due to the replacement of cysteine at position 104 in framework region 3 with glycine and *IGKV3-7* has a mutation in an acceptor splice site (Pargent et al., 1991). The reason for stringent selection against *IGKV5-2* is less clear. The *IGKV* genes that are selected into the developing repertoire tend to be the most commonly used in previous studies including *IGKV1-39*, *IGKV3-20*, *IGKV3-11*, *IGKV1-5* (Collins et al., 2008; Foster et al., 1997; Jackson et al., 2013).

Although *IGHV* undoubtedly contributes to auto-antigen recognition in SLE, frequently emphasis of the role of *IGKV* gene segment usage has been given (Ibrahim et al., 1995; Manheimer-Lory et al., 1995a; Suzuki et al., 1996). The gene *IGKV4-1* was of considerable interest and had a selection profile quite different to any other gene in health. The high frequency of rearrangement *of IGKV4-1* may be related to its proximity to IGKJ. Germline transcripts of *IGKV4-1* have been found in pre B cells and foetal bone marrow cells, showing that this gene segment is readily available (Martin et al., 1991). In DNA from healthy donors, rearrangements of *IGKV4-1* were the most frequent rearrangements detected, but they were significantly less frequent in cDNA. This suggests strongly that

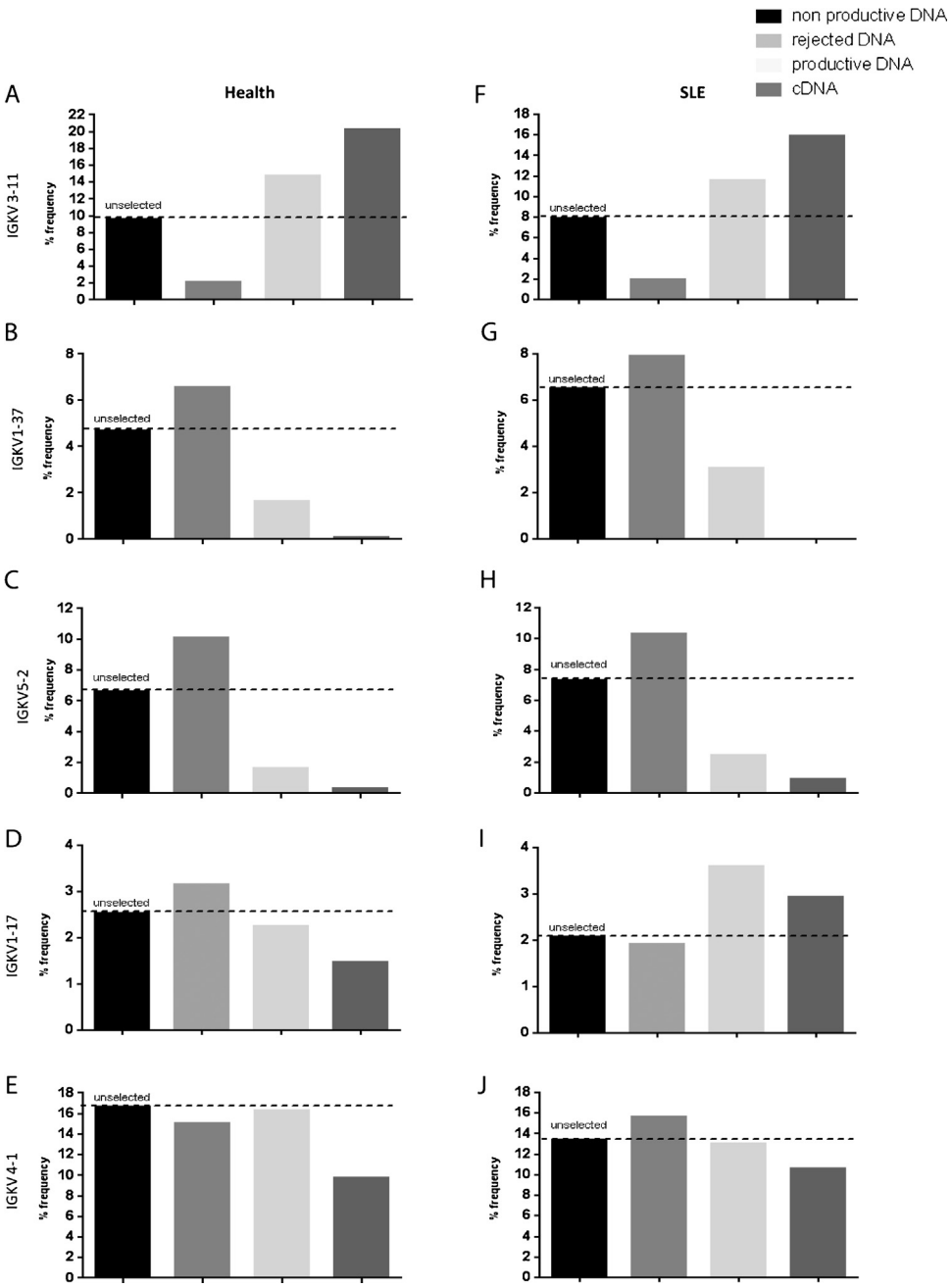


Fig. 4. Selection profile of five *IGKV* genes in health and SLE. Comparison of relative frequencies of rearrangements from four groups of PCR amplicons: non-productive rearrangements of IGK expressing B cells, rejected DNA (=productive DNA rearrangements in IGL expressing B cells), productive DNA rearrangements of IGK expressing B cells and expressed cDNA of IGK expressing mature naïve B cells were compared. Selection profiles: selection for *IGKV3-11* in (A) health and (F) SLE. Selection against *IGKV1-37* in (B) health and (G) SLE. Selection against *IGKV5-2* in (C) health and (H) SLE. Selection against *IGKV1-17* in expressed cDNA repertoire in (D) health and apparently less selected against in SLE (I). Selection against *IGKV4-1* in expressed cDNA repertoire in (E) health and apparently less stringently selected against in SLE (J).

IGKV4-1 is selected against in health. However, rearrangements of *IGKV4-1* were neither more nor less abundant in the productively rearranged DNA from IGK or IGL expressing B cells, or when compared to non-productive DNA rearrangements. This is in agreement with a study comparing productively and non-productively rearrangements in genomic DNA of CD19^{+} sorted cells in health (Dorner et al., 1998). The unusual features of *IGKV4-1* observed are not likely to be purely consequence its inverse orientation. *IGKV5-2* is in the same orientation as *IGKV4-1* and is also a J-proximal gene adjacent to *IGKV4*-1 on unrearranged chromosome 2. However, there is a strong imprint of selection against the usage of *IGKV5-2* throughout in our model, which is similar to genes known to be functionally defective.

The higher frequency of rearrangements of *IGKV4-1* in DNA from IGK expressing cells compared to cDNA in health was consistent in all six healthy blood donors and was statistically significant when compared as a population. However, this was not observed in two of three patients with SLE where the frequencies of *IGKV4-1* in the productive rearrangements from DNA and cDNA of IGK expressing B cells were approximately the same. Differences in coding region functionality and gene editing for a particular gene segment were observed when comparing SLE patients with varying clinical features (Suzuki et al., 1996). Over-expression of *IGKV4-1* has been observed in some patients with SLE (Dorner et al., 1998; Yurasov et al., 2005a, 2005b), and it is possible that the aberrant profiles of *IGKV*-J rearrangements in SLE may be due to defects in regulating

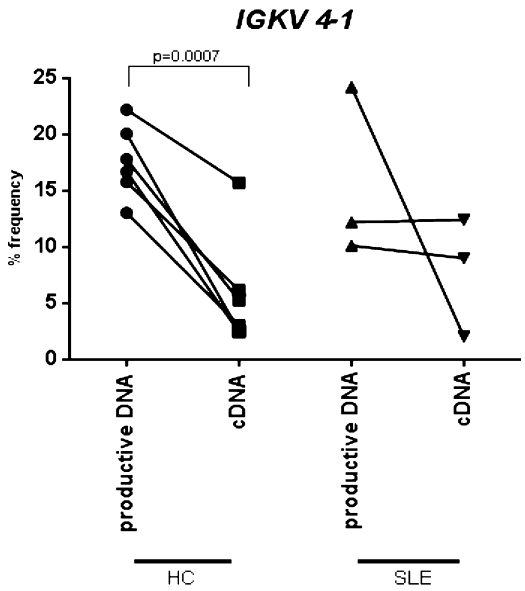


Fig. 5. Selection profile of *IGKV4-1* gene segment. Selection profile of *IGKV4-1* in the expressed cDNA repertoire of six healthy individuals and three SLE patients from DNA and cDNA of IGK expressing mature naïve B cells. Selection against expression of *IGKV4-1* in health but not in SLE. Statistical analysis in was performed using Student's *t*-test, $p < 0.05$. *p*-Values are indicated as $^{*}p < 0.05$.

IGK locus inactivation. Several studies have been undertaken investigating the possibility of aberrant receptor editing in patients with SLE (Bensimon et al., 1994). Suzuki et al. (1996) analysed receptor editing for *IGVK1-17* segment (A30) in health and patients with SLE with and without nephritis. It has been shown that *IGKV1-17* is frequently used in cationic anti-DNA antibodies in SLE patient with nephritis, suggesting failure in secondary gene rearrangement for these groups of patients. Consistent with this we observed more abundant *IGKV1-17* in DNA than cDNA in health, but less so in SLE.

Dorner et al. (1998) showed in their study by comparing *IGKV-J* gene distribution that receptor editing was not defective in early, untreated SLE patients when compared to healthy individuals. Consistent with this study, there was no difference between the rearrangements of *IGKV* genes between health and SLE when the samples from DNA were compared across this study. Reduced frequency of rearrangement of the KDE has been reported in SLE as a surrogate measurement for receptor editing (Panigrahi et al., 2008). It is also possible, that inefficient rearrangement of the KDE could permit expression of gene rearrangements in SLE that would not be expressed in health.

The relative use of up and down stream *IGKJ* segments has been used to indicate receptor editing or revision in previous studies (Dorner et al., 1998; Yurasov et al., 2006). However, this is difficult to interpret at the *IGK* locus due to the different orientations of gene segments. We observe a consistently higher frequency of *IGKJ5* and lower frequency of *IGKJ1* in DNA compared to cDNA in health and SLE which is consistent with previous studies (Dorner et al., 1998; Manheimer-Lory et al., 1995a, 1995b). However, in a study from

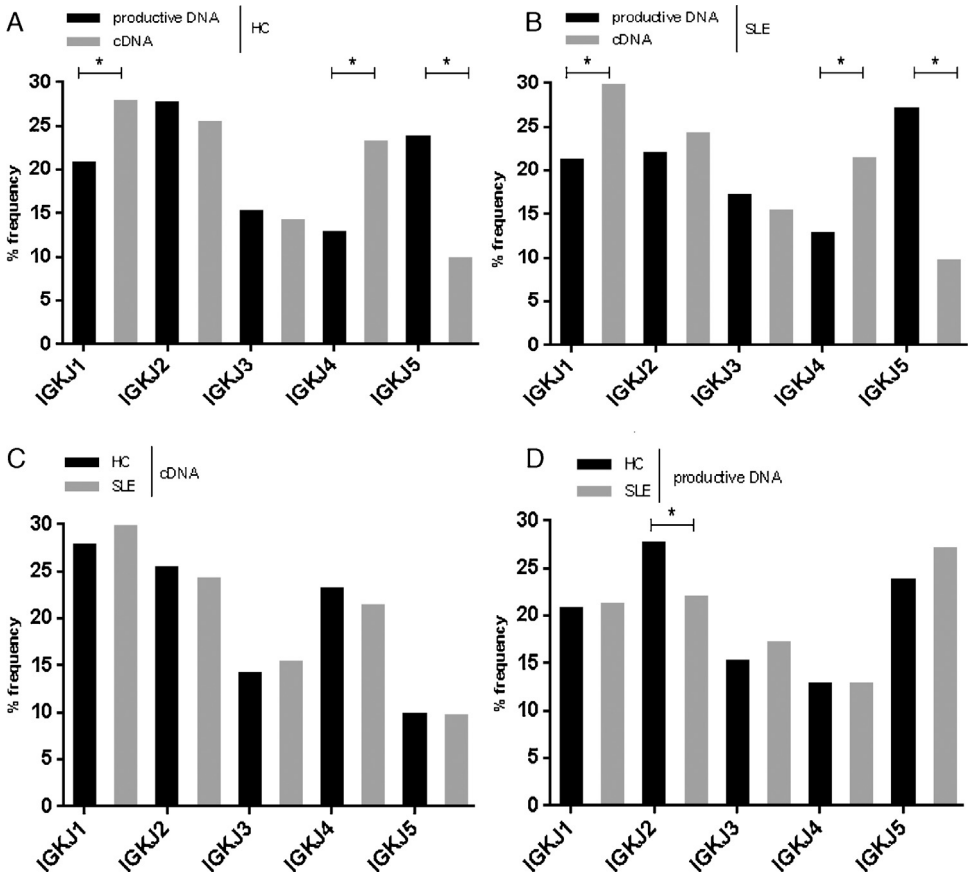


Fig. 6. Comparison of relative frequencies of rearrangements of *IGKJ* gene segments. Data from PCR amplicons from productive DNA and expressed cDNA of IGK expressing mature naïve B cells from health and SLE (A and B). Selection bias is apparent for *IGKJ1*, *IGKJ2* and *IGKJ3* for both health and SLE. Comparison of relative frequencies of rearrangements of *IGKJ* gene segments from PCR amplicons from productive DNA of health and SLE (C) and expressed cDNA from health and SLE (D) of IGK expressing mature naïve B cells showing a similar profile. Statistical analysis was performed using χ^2 analysis adjusted using the Bonferroni correction. Significance is indicated as $^{*}p < 0.05$.

Table 1
Absolute numbers of unique IGKV sequences. A total of 20,160 unique gene rearrangements of IGK and IGL expressing cells were analysed. Four groups of PCR amplicons from DNA and one from cDNA have been included.

Number of unique rearrangements of IGKV genes analysed								
	From sorted IGK expressing mature naïve B cells				From sorted IGL expressing mature naïve B cells			
	1	2	3		3	4		
Group of sequences	cDNA	Productively rearranged DNA	Non-productively rearranged DNA		Non-productively rearranged DNA	Productively rearranged DNA = rejected		
Healthy	5643	1908	1542		4049	1401		
SLE	2214	918	730		1338	417		

Yurasov et al. (2006) bias was observed towards the usage of *IGKJ* distal segments in SLE patients with an active disease profile. This did not differ between any of the gene segments analysed and was broadly similar between health and SLE.

The differences between rearrangements of *IGKV* in amplified DNA and cDNA that we observed in this study are not the consequence of PCR primers used to generate them, as primer mixes used for V segments were identical. The profile of *IGKJ* segments was not different when different *IGKV* segments were studied separately and any biases in *IGKJ* binding would not have resulted in biases in *IGKV*. A shortcoming of the study is the unequal numbers of rearrangements analysed. However, the data in general was consistent across all six healthy donors studied. The number of sequences and also the number of patients with SLE studied was relatively low and it will be important to follow the concepts described here using other models in the future.

In conclusion we found differences in the expressed *IGKV* B cell repertoire between health and SLE that were not apparent in DNA alone. Gene segments that appear to be selected against are more abundant in cDNA in patients with SLE than in health. Our data implies that selection of IGK genes into the B cell repertoire is regulated at different stages of B cell development. Choices are made during B cell development, e.g., at the stage when cell fate decisions are made to either edit *IGK* or rearrange the *IGL* locus. These appear to be similar between health and SLE since the imprints of these processes are apparent in the DNA rearrangements analysed. Our data suggest that the activity of the KDE itself can contribute to the composition of the expressed repertoire and that this is inefficient in SLE. There are numerous genetic polymorphisms associated with SLE that relate to B cell signalling including those in Blk and BANK1 that could be associated with B cell activity and the regulation of gene rearrangements including KDE (Castillejo-Lopez et al., 2012; Deng and Tsao, 2010; Jarvinen et al., 2012). Our data suggest that factors that regulate expression of productive rearrangements of *IGKV-J*, that have been selected against, such as rearrangement of the KDE or the intronic RSS, may be involved in shaping the expressed repertoire of *IGKV* genes, and that this process may be aberrant in some way in SLE.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

Acknowledgements

This study was supported by the St Thomas' Wegener's Trust and the Oliver Bird Rheumatism Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.molimm.2015.01.017.

References

Bensimon, C., Chastagner, P., Zouali, M., 1994. Human lupus anti-DNA autoantibodies undergo essentially primary V kappa gene rearrangements. EMBO J. 13, 2951–2962.

Brauninger, A., Goossens, T., Rajewsky, K., Kuppers, R., 2001. Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human. Eur. J. Immunol. 31, 3631–3637.

Castillejo-Lopez, C., Delgado-Vega, A.M., Wojcik, J., Kozyrev, S.V., Thavathiru, E., Wu, Y.Y., Sanchez, E., Pollmann, D., Lopez-Egido, J.R., Fineschi, S., Dominguez, N., Lu, R., James, J.A., Merrill, J.T., Kelly, J.A., Kaufman, K.M., Moser, K.L., Gilkeson, G., Frostegard, J., Pons-Estel, B.A., D'Alfonso, S., Witte, T., Callejas, J.L., Harley, J.B., Gaffney, P.M., Martin, J., Guthridge, J.M., Alarcon-Riquelme, M.E., 2012. Genetic

and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. Ann. Rheum. Dis. 71, 136–142.
Collins, A.M., Wang, Y., Singh, V., Yu, P., Jackson, K.J., Sewell, W.A., 2008. The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate. Immunogenetics 60, 669–676.
Deng, Y., Tsao, B.P., 2010. Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nat. Rev. Rheumatol. 6, 683–692.
Dorner, T., Foster, S.J., Farner, N.L., Lipsky, P.E., 1998. Immunoglobulin kappa chain receptor editing in systemic lupus erythematosus. J. Clin. Investig. 102, 688–694.
Foster, S.J., Brezinschek, H.P., Brezinschek, R.I., Lipsky, P.E., 1997. Molecular mechanisms and selective influences that shape the kappa gene repertoire of IgM+ B cells. J. Clin. Investig. 99, 1614–1627.
Goossens, T., Brauninger, A., Klein, U., Kuppers, R., Rajewsky, K., 2001. Receptor revision plays no major role in shaping the receptor repertoire of human memory B cells after the onset of somatic hypermutation. Eur. J. Immunol. 31, 3638–3648.
Hardy, R.R., Hayakawa, K., 2001. B cell development pathways. Annu. Rev. Immunol. 19, 595–621.
http://www.imgt.org/IMGTrepertoire/, http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=locus&species=human&group=IGK
Ibrahim, S.M., Weigert, M., Basu, C., Erikson, J., Radic, M.Z., 1995. Light chain contribution to specificity in anti-DNA antibodies. J. Immunol. (Baltimore, Md.: 1950) 155, 3223–3233.
Jackson, K.J., Kidd, M.J., Wang, Y., Collins, A.M., 2013. The shape of the lymphocyte receptor repertoire: lessons from the B cell receptor. Front. Immunol. 4, 263.
Jarvinen, T.M., Hellquist, A., Zucchelli, M., Koskenmies, S., Panelius, J., Hasan, T., Julkunen, H., D'Amato, M., Kere, J., 2012. Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population. Rheumatology (Oxf., Engl.) 51, 87–92.
Manheimer-Lory, A., Irigoyen, M., Gaynor, B., Monhian, R., Splaver, A., Diamond, B., 1995a. In: Casali, P., Silberstein, L.E. (Eds.), Analysis of V kappa I and V lambda II light chain genes in the expressed B-cell repertoire, Vol. 764. New York Academic Sciences, New York, pp. 301–311.
Manheimer-Lory, A., Monhian, R., Splaver, A., Gaynor, B., Diamond, B., 1995b. Analysis of the V kappa I family: germline genes from an SLE patient and expressed autoantibodies. Autoimmunity 20, 259–265.
Martin, D., Huang, R.Q., LeBien, T., Van Ness, B., 1991. Induced rearrangement of kappa genes in the BLIN-1 human pre-B cell line correlates with germline J-C kappa and V kappa transcription. J. Exp. Med. 173, 639–645.
Panigrahi, A.K., Goodman, N.G., Eisenberg, R.A., Rickels, M.R., Naji, A., Luning Prak, E.T., 2008. RS rearrangement frequency as a marker of receptor editing in lupus and type 1 diabetes. J. Exp. Med. 205, 2985–2994.
Pargent, W., Meindl, A., Thiebe, R., Mitzel, S., Zachau, H.G., 1991. The human immunoglobulin kappa locus. Characterization of the duplicated O regions. Eur. J. Immunol. 21, 1821–1827.
Santos, P., Arumemi, F., Park, K.S., Borghesi, L., Milcarek, C., 2011. Transcriptional and epigenetic regulation of B cell development. Immunol. Res. 50, 105–112.
Schoettler, N., Ni, D., Weigert, M., 2012. B cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients. Mol. Immunol. 51, 273–282.
Steinsbø, Ø., Dunand, C.J.H., Huang, M., Mesin, L., Salgado-Ferrer, M., Lundin, K.E.A., Jahnsen, J., Wilson, P.C., Sollid, L.M., 2014. Restricted VH/VL usage and limited mutations in gluten-specific IgA of coeliac disease lesion plasma cells. Nat. Commun., 5.
Suzuki, N., Harada, T., Mihara, S., Sakane, T., 1996. Characterization of a germline Vk gene encoding cationic anti-DNA antibody and role of receptor editing for development of the autoantibody in patients with systemic lupus erythematosus. J. Clin. Investig. 98, 1843–1850.
Wardemann, H., Yurasov, S., Schaefer, A., Young, J.W., Meffre, E., Nussenzweig, M.C., 2003. Predominant autoantibody production by early human B cell precursors. Science 301, 1374–1377.
Woodward, E.J., Thomas, J.W., 2005. Multiple germline kappa light chains generate anti-insulin B cells in nonobese diabetic mice. J. Immunol. (Baltimore, Md.: 1950) 175, 1073–1079.
Wu, Y.C., Kipling, D., Leong, H.S., Martin, V., Ademokun, A.A., Dunn-Walters, D.K., 2010. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood 116, 1070–1078.
Yurasov, S., Hammersen, J., Tiller, T., Tsuiji, M., Wardemann, H., 2005a. B-cell tolerance checkpoints in healthy humans and patients with systemic lupus erythematosus. Ann. N.Y. Acad. Sci. 1062, 165–174.
Yurasov, S., Tiller, T., Tsuiji, M., Velinzon, K., Pascual, V., Wardemann, H., Nussenzweig, M.C., 2006. Persistent expression of autoantibodies in SLE patients in remission. J. Exp. Med. 203, 2255–2261.
Yurasov, S., Wardemann, H., Hammersen, J., Tsuiji, M., Meffre, E., Pascual, V., Nussenzweig, M.C., 2005b. Defective B cell tolerance checkpoints in systemic lupus erythematosus. J. Exp. Med. 201, 703–711.